

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 17, 2001, 08:14:16 ; Search time 1543.18 Seconds
(without alignments)
3822.906 Million cell updates/sec

Title: US-09-462-625-1

Perfect score: 549

Sequence: 1 atgtgtttgcctgtgctct.....gggaacactaccgagagtga 549

Scoring table: OLIGO_NUC

Gapop 60.0 , Gapext 60.0

Searched: 11351937 seqs, 537289281 residues

Word size : 0

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

EST:*

- 1: em_estfun:*
- 2: em_esthum:*
- 3: em_estin:*
- 4: em_estpl:*
- 5: em_estom:*
- 6: em_estba:*
- 7: em_estro:*
- 8: em_estov:*
- 9: em_hic:*
- 10: gb_estl:*
- 11: gb_est2:*
- 12: gb_hic:*
- 13: gb_gss:*
- 14: em_gss_fun:*
- 15: em_gss_hum:*
- 16: em_gss_inv:*
- 17: em_gss_pln:*
- 18: em_gss_pro:*
- 19: em_gss_rod:*
- 20: em_gss_vrt:*
- 21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	549	100.0	682	12 AK008335	Mus muscu
2	498	90.7	589	11 BG174272	602334571
3	498	90.7	637	11 B1149595	602848402
4	498	90.7	688	11 BG244455	602336590
5	489	89.1	610	11 B1154844	602902826
6	475	86.5	600	11 BG871384	602790566
7	466	84.9	645	11 BG975104	602843369
8	447	81.4	650	11 B1155774	602904343
9	441	80.3	1045	11 B1155774	601669059
10	416	75.8	842	11 B1409815	602961906
11	394	71.8	518	10 B135520	u952e11.y
12	379	69.0	575	10 A1507116	vj62109.x

13	377	68.7	706	11 BF163190	601771917
14	370	67.4	619	11 BF302505	602031534
15	354	64.5	384	10 BE654595	UI-M-AJ1-
16	339	61.7	614	10 BE199698	u952e11.x
17	284	51.7	451	11 BF151994	u222h10.y
18	278	50.6	548	10 AA597240	u222h10.y
19	272	49.5	528	10 AA734893	u222h10.y
20	270	49.2	534	10 AA734805	u222h10.y
21	268	48.8	510	11 BF152060	u222h10.y
22	221	40.3	547	10 AA689693	u222h10.y
23	221	40.3	591	10 AA238564	u222h10.y
24	221	40.3	601	10 AA238752	u222h10.y
25	192	35.0	548	10 A1585767	u222h10.y
26	172	31.3	384	10 AA689633	u222h10.y
27	149	27.1	665	11 BG969181	u222h10.y
28	142	25.9	564	10 AA499942	u222h10.y
29	125	22.8	573	10 AA543570	u222h10.y
30	118	21.5	571	11 BG963242	u222h10.y
31	112	20.4	274	11 B1159353	u222h10.y
32	107	19.5	305	10 AA930493	u222h10.y
33	107	19.5	316	10 AA930504	u222h10.y
34	104	18.9	322	10 A1849253	u222h10.y
35	99	18.0	195	11 BF227350	u222h10.y
36	98	17.9	427	10 AA734995	u222h10.y
37	97	17.7	329	10 AA616255	u222h10.y
38	80	14.6	982	11 BF160271	u222h10.y
39	70	12.8	235	10 BB581812	u222h10.y
40	59	10.7	281	10 AA691322	u222h10.y
41	55	10.0	277	10 AY085455	u222h10.y
42	48	8.7	234	10 A1153056	u222h10.y
43	42	7.7	150	11 D19431	u222h10.y
44	41	7.5	284	10 AV092014	u222h10.y
45	35	6.4	226	10 AV082572	u222h10.y

ALIGNMENTS

RESULT 1

AK008335

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

AK008335 682 bp mRNA HTC 05-JUL-2001
Mus musculus adult male small intestine cDNA, RIKEN full-length
enriched library, clone:2010107E11, full insert sequence.

AK008335

AK008335.1 GI:12842459

CAP trapper.

Mus musculus (strain:C57BL/6J) adult male small intestine cDNA to

mRNA, clone:lib:RIKEN full-length enriched mouse cDNA library

clone:2010107E11.

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 682)

Carninci, P. and Hayashizaki, Y.

High-efficiency full-length cDNA cloning

Methods in enzymology. 303, 19-44 (1999)

99279253

10349636

2 (bases 1 to 682)

Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,

Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.

Normalization and subtraction of cap-trapper-selected cDNAs to

prepare full-length cDNA libraries for rapid discovery of new genes

Genome research. 10 (10), 1617-1630 (2000)

20499374

11042159

3 (bases 1 to 682)

Shibata, K., Itoh, M., Aizawa, K., Katsunai, T., Tashiro, H., Itoh, M.,

Konno, H., Akiyama, J., Nishi, K., Katsunai, T., Tashiro, H., Itoh, M.,

Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,

Yanamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,

Fujitake, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M.,

Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsubara, S., Kawai, J.,


```

/lab_host="DH10B"
/notes=:organ; mammary; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert 2.5 kb. Library constructed by Life
Technologies, catalog # 12018-016. Investigators providing
samples: Lothar Hennighausen/Priscilla Furth, NIH
Reference for transgenic model: Li et al., Cell Growth and
Differentiation 7, 3-11 (1996). Note: this is a NCI-CGAP
Library."
202 c 176 g 133 t

```

```

CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM1094 row: f column: 12
High quality sequence stop: 637.
Location/Qualifiers
1. .650
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:5033867"
/clone_lib="NIH_CGAP_Mam3"
/tissue_type="tumor, gross tissue"
/lab_host="DH10B"
/note="Organ: mammary; Vector: pCMV-SPORT6; Site.1: NotI;
Site.2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert 2 kb. Library constructed by Life
Technologies, catalog #12017-018. Investigators providing
samples: Lothar Hennighausen/Chu-Xia Deng, NIH Reference
for transgenic model: Xu et al., Nature Genetics 22, 37-43
(1999). Note: this is a NCI-CGAP Library."
137 a 179 q 135 t
BASE COUNT

```



```
Db 1 |||||TAGCGCCTGGGGCACCAGTTCGCTAGCTGGTGATCTCACACAGCGGCA 60
Qy 155 9cttctgcaacagcccgactctgtgaacagcagcccgcaatgtgcagcattaccaca 214
Db 61 |||||CTGCAACAGCCCGGACTCCTGTGAACAGCAGGCCCGCAATGTGAGCATTTACCACA 120
Qy 215 agaatsagctgggtgctgtagcttagcctacaacttcttattggagagacggtcatg 274
Db 121 AGAATGAGCTGGGCTGGTGCATGTAGCTACAACTTCCTTATTGGAGGACGGTCATG 180
Qy 275 tctatgaagccgaggtggaacatcaaggtgaccacacagggcccatcttggaatccca 334
Db 181 TCTATGAAGCCGAGGCTGGAACTCAAGGGTGACCAACAGGGGCCCATCTGGAATCCCA 240
Qy 335 tgtctattggcatcaccttctatggggaacttcattggaccgggtaccgcgaagcgggccc 394
Db 241 TGTCTATTGGCATCACCTTTCATGGGGAACCTTCATGGACCGGGGTACCCGCAAGCGGGCCC 300
Qy 395 tccgtgctgcccctaaattcttctggaatgtgggtgtctctgggggttctctgagat 448
Db 301 TCCGTGCTGCCCTAAATCTTCTTGGAAATGGGGTGCTCTCGGGGCTTCTCTGAGAT 354
```

Search completed: December 17, 2001, 10:01:19
Job time: 6423 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 17, 2001, 07:47:55 ; Search time 1549.8 Seconds
(without alignments)
3806.577 Million cell updates/sec

Title: US-09-462-625-1
Perfect score: 549
Sequence: 1 atgtgttgctgtgtct.....gggaacactaccgagagtga 549

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

EST:*

- 1: em_estfun:**
- 2: em_esthum:**
- 3: em_estim:**
- 4: em_estom:**
- 5: em_estpl:**
- 6: em_estba:**
- 7: em_estro:**
- 8: em_estov:**
- 9: em_htc:**
- 10: gb_est1:**
- 11: gb_est2:**
- 12: gb_htc:**
- 13: gb_gss:**
- 14: em_gss_fun:**
- 15: em_gss_hum:**
- 16: em_gss_inv:**
- 17: em_gss_pln:**
- 18: em_gss_pro:**
- 19: em_gss_rod:**
- 20: em_gss_vrt:**
- 21: em_gss_other:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	549	100.0	682	12	AK008335
2	547.4	99.7	589	11	BG174272
3	547.4	99.7	637	11	B1149595
4	547.4	99.7	688	11	BG244455
5	545.8	99.4	650	11	B1155774
6	539.8	98.3	610	11	B1154844
7	536.4	97.5	600	11	BG871384
8	535.4	97.7	706	11	BF163190
9	532.2	96.9	645	11	BG975104
10	532.2	96.9	842	11	B1409815
11	528.4	96.2	619	11	BF302505
12	525.4	95.7	1045	11	BE913706

ALIGNMENTS

RESULT 1

AK008335 682 bp mRNA HTC 05-JUL-2001
Mus musculus adult male small intestine cDNA, RIKEN full-length
enriched library, clone:2010107E11, full insert sequence.

AK008335 1 GI:12842459
CAP trapper.

Mus musculus (strain:C57BL/6J) adult male small intestine cDNA to
mRNA, clone_lib:RIKEN full-length enriched mouse cDNA library
clone:2010107E11.

Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 682)
Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning
Methods in enzymology. 303, 19-44 (1999)

2 (bases 1 to 682)
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
Genome research. 10 (10), 1617-1630 (2000)

20499374
11042159
3 (bases 1 to 682)
Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
Konno, H., Akiyama, J., Nishi, K., Kitsuai, T., Tashiro, H., Itoh, M.,
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M.,
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J.,

13 516.6 94.1 601 10 AA238752
14 505 92.0 614 10 BE196698
15 501.8 91.4 591 10 AA238564
16 465.6 84.8 575 10 AT507116
17 464.6 84.6 528 10 AA734993
18 437.8 79.7 518 10 BE135520
19 422.4 76.9 547 10 AA689693
20 412.8 75.2 627 10 AA228200
21 408.4 74.4 534 10 AA734805
22 403.6 73.5 548 10 AA597240
23 400.8 73.0 548 10 AT585767
24 392.8 71.5 573 10 AA543570
25 386.6 70.4 564 10 AA499942
26 381.4 69.5 384 10 BE654595
27 361.8 65.9 665 11 BG69181
28 354.8 64.6 510 11 BF152060
29 349.2 63.6 384 10 AA689633
30 335.2 61.1 451 11 BF151994
31 271 49.4 427 10 AA734995
32 263.8 48.1 555 11 BF076851
33 263.2 47.9 575 10 AA467554
34 257.6 46.9 472 11 BI291182
35 257.6 46.9 506 10 AA875213
36 251.2 45.8 504 11 BF548605
37 239 43.5 520 10 AW076051
38 237 43.2 384 10 AA015317
39 221.6 40.4 329 10 AA616255
40 202.6 36.9 322 10 A1849253
41 185.4 33.8 305 10 AA930493
42 171.4 31.2 277 10 AV085455
43 171.4 31.2 281 10 AA691322
44 169.8 30.9 316 10 AA930504
45 167.4 30.5 571 11 BG963242

<hr/>					
ACCESSION	BC244455				
VERSION	BC244455.1	GI:12754270			
KEYWORDS	EST.				
SOURCE	house mouse.				
ORGANISM	Mus musculus				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
AUTHORS	Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus;				
TITLE	1 (bases 1 to 688)				
JOURNAL	NIH-MGC http://mgc.nci.nih.gov/.				
COMMENT	National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999) Contact: Robert Strausberg, Ph.D. Email: cygabs@remail.nih.gov Tissue Procurement: Gilbert Smith, Ph.D. cDNA Library Preparation: Life Technologies, Inc. cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Plate: LLAM10326 row: c column: 24 High quality sequence stop: 587.				
FEATURES	Location/Qualifiers				
source	1..688				
	/organism="Mus musculus"				
	/strain="FVB/N"				
	/db_xref="taxon:10090"				
	/clone="IMAGE:4485071"				
	/clone_lib="NCI-CGAP_Maml"				
	/tissue_type="tumor, biopsy sample"				
	/dev_stage="3 months, virgin"				
	/lab_host="DH10B"				
	/note="Organ: mammary; Vector: pCMV-SPORT6; Site1: Salt; Site2: NotI; Cloned unidirectionally. Primer: Oligo dT Library constructed by Life Technologies. Investigator providing samples: Gilbert Smith, NIH"				
BASE COUNT	157 a 209 c 180 g 142 t				
ORIGIN					
Query Match	99.7%; Score 547.4; DB 11; Length 688;				
Best Local Similarity	99.8%; Pred. No. 1.4e-134;				
Matches	548; Conservative 0; Mismatches 1; Indels 0; Gaps 0				
Qy	1 atgttgttcctgtcctcccttgccctcctggctggcgcaacctctgcagtttaac 60				
Dd					
	19 ATGTTGTTCCTTGCTCTCCTTGGCCCTCTCGGCTGCGCAACCTCTCTGCAGTTTCATC 78				
Qy	61 gtccccgcagtgagtggaggggcccctgccatccgagtgctctagccgctggggcaccca 120				
Dd					
	79 GTCCCCGCGAGTGAGTGGAGGGGCCCTGCCATCCGAGTGTCTAGCGGCTGGGGCACCCA 138				
Qy	121 gttcgtacgtgtgatctcaacacagccggcagctcttgcaacagccggactcctgt 180				
Dd					
	139 GTTCGCTACGTGTGTGATCTACACACAGCCGCAGCTTCTGCAACAGCCGCGACTCTCTGT 198				
Qy	181 gaacagcagggccgcgaattgtagcattaccacaagaatgagctggctggcgcatga 240				
Dd					
	199 GAACAGAGGCCCGCAATGTGCAGCATTTACCACAGAATGACCTGGGTGTGTCGATGTA 258				
Qy	241 gcctacaacttccttatgttgaggaggacgtcatgtcttatgaaggccgaggtcgaaacatc 300				
Dd					
	259 GCCTACAACTTCCTATTGGAGAGGACGGTCTATGTCATGAAGCGCGAGGCTGGAAACATC 318				
Qy	301 aagggtgaccacacagggcccatctggaatcccatgtcttatggcatcaccttcaatgggg 360				
Dd					
	319 AAGGGTGACCACACAGGGCCCATCTGGAAATCCCATGTCTATTGGCATCACTTTCATGGG 378				
Qy	361 aacttcatgacccgggtaccccgcaagcgggcccctcgctgctgcctaaactcttcctgaa 420				
Dd					
	379 AACTTCATGACACCGGGTACCCGCAAGCGGGCCCTCCGTGCTGCCCTTAATCTTCGAAA 438				
Qy	421 tgtgggggtgtctcggggctcttccttgagatccaactatgaagtcaaggacacccgggatgtg 480				

```

QY 181 gaacagcagccgcaatgtgcagcattaccacaaagaatagctggctgtgcatgta 240
Db 251 GAACAGCAGCCCGCAATGTCAGCATTAACCAAGAATGAGCTGGCTGGTGGATGA 310
QY 241 gctacaacttccttattgagagcagtcgtctatgataagcagcagctggaacatc 300
Db 311 GCTACAACCTCTTATTGGAGAGGAGCGGTCATGTATGAAGCCGAGGCTGGAACATC 370
QY 301 aagggtagcacacagagcccatctggaatcccatctctattggcatcaccttcattgggg 360
Db 371 AAGGGTAGCACACAGAGGCCCATCTGGAATCCCATGCTATTGGCATCACCCTTCATGGGG 430
QY 361 aacttcattgagcgggtaccgcgaagcgggcccctccgtgctgccttaaatctctggaa 420
Db 431 AACTTCATGACCGGTACCCGCAAGACGGGCCCTCCGTGCTGCCCTACATCTTCGGAA 490
QY 421 tctggggtctcggggctctctgagatccaaactatgaagtcaagacacccgggatgtg 480
Db 491 TGTGGGCTCTCGGGGCTCTCTTGAGATCCAACTATGAAGTCAAGGACACCCGGGATGTG 550
QY 481 caaagcactctctccagtgaccactctatcagggtcatccaaagctgggaacactac 540
Db 551 CAAAGCAGCTCTCTCCAGGTGACCAACTATATCAGGTATCCAAAGCTGGGAACACTAC 610
QY 541 cgagagtga 549
Db 611 CGAGAGTGA 619

RESULT 6
BI154844 610 bp mRNA EST 05-JUL-2001
LOCUS 602902826F1 NIH_CGAP_Mam3 Mus musculus cDNA clone IMAGE:5032278 5',
DEFINITION mRNA sequence.
ACCESSION BI154844
VERSION BI154844.1 GI:14614845
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE NIH-MGC http://mgc.nci.nih.gov/.
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Tissue Procurement: Lothar Hennighausen Ph.D., Chu-Xia Deng Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: LLAM1090 row: d column: 07
High quality sequence stop: 610.
FEATURES
Location/Qualifiers
1..610
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:5032278"
/collection="NIH_CGAP_Mam3"
/tissue.type="tumor, gross tissue"
/lab_host="DH10B"
/Note="Organ: mammary; Vector: pCMV-SPORT6; Site: 1: NotI;
Site: 2: SalI; Cloned unidirectionally. Primer: Oligo dr.
Average insert 2 kb. Library constructed by Life
Technologies, catalog #12017-018. Investigators providing
samples: Lothar Hennighausen/Chu-Xia Deng, NIH Reference
for transgenic model: Xu et al., Nature Genetics 22, 37-43
(1999). Note: this is a NIH_CGAP Library."
BASE COUNT 128 a 179 c 173 g 130 t

```

ORIGIN

Query Match 98.3%; Score 539.8; DB 11; Length 610;
 Best Local Similarity 99.8%; Pred. No. 1.4e-132;
 Matches 541; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

QY 7 ttgctgtgtctctccttgcctctctgggtctggcaacctctcagttctcatgtgccc 66
Db 38 TTGGCTGTGCTCTCTCTTGGCTCTGGGTCTGGCAACCTCTCGAGTTTCTATCGTGCCC 97
QY 67 cgagtgagtgaggggccctgccaatccgagtgctctagccgctggggacccagttgac 126
Db 98 CGCAGTGAAGTGGAGGGCCCTGCCATCGAGTGTCTAGCCGCTGGGGCACCAAGTTCGC 157
QY 127 tagtggtgattcacaacacagcggcagctctgcaacagccgagactctgtgaacag 186
Db 158 TAGGTGTGATCTCACACAGCGCGGAGCTCTGTCAACGCCGGACTCTCTGTGAACAG 217
QY 187 caggccgcgaatgtgcagcattaccacaagaatgagctgggtggtgctgtagcttac 246
Db 218 CAGCCCGCAATGTCAGCATTAACCAAGAATGAGCTGGCTGGTGGATGTAGCCTAC 277
QY 247 aacttcattgagagcagtgctctatgataagcagcagctgggaacatcaaggt 306
Db 278 AACTTCCTTATTGGAGAGGAGCGGTCTATGTAAGCCGAGGCTGGAACATCAAGGT 337
QY 307 gaccacacagggcccatctggaatcccatctctattggcatcaccttcattgggaacttc 366
Db 338 GACCACACAGGGGCCCATCTGGAATCCCATGCTATTGGCATCACCCTTCATGGGAATTC 397
QY 367 atggaccgggtaccgcgaagcgggcccctccgtgctgccttaaatctctggaatgtgg 426
Db 398 ATGACCGGGTACCCGCAAGAGCGGCCCTCCGTGCTGCCCTAAATCTCTCGAATGTGGG 457
QY 427 gtgtctcggggctctctgagatccaaactatgaagtcgaagacacccgggatgtgcaagc 486
Db 458 GTGTCTCGGGGCTCTTGTGAGATCCAACTATGAAGTCAAGGACACCCGGGATGTCAAAAGC 517
QY 487 actctctccagtgaccacactctatcagggtcatccaaagctgggaacacatccagag 546
Db 518 ACTCTCTCTCCAGGTGACCACTCTATCAGGTCTATCAAGTCTCAAAAGCTGGGAACACTACCGAGAG 577
QY 547 tga 549
Db 578 TGA 580

```

RESULT 7

```

BI154844 600 bp mRNA EST 29-MAY-2001
LOCUS 602790566F1 NCI_CGAP_SG2 Mus musculus cDNA clone IMAGE:4921697 5',
DEFINITION mRNA sequence.
ACCESSION BG871384
VERSION BG871384.1 GI:14221924
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE NIH-MGC http://mgc.nci.nih.gov/.
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov

```



```

Db 541 CAAGCACTCTCTCCAGGTGACCAACTCTATCAGGTATCCAAAGCTGGGAACACTAC 600
QY 541 cgagagtga 549
Db 601 CGAGAGTGA 609

RESULT 9
LOCUS BG975104 645 bp mRNA 12-JUN-2001
DEFINITION 602843369F1 NIH_CGAP_Mam4 Mus musculus cDNA clone IMAGE:4979057 5',
rna sequence.
ACCESSION BG975104
VERSION BG975104.1 GI:14362741
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 645)
TITLE NIH-MGC http://mgc.nci.nih.gov/.
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: Lothar Hennighausen Ph.D., Priscilla Furth
Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM10976 row: j column: 18
High quality sequence start: 3
High quality sequence stop: 634.
Location/Qualifiers
1. 645
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:4979057"
/clone_lib="NIH_CGAP_Mam4"
/tissue_type="tumor, gross tissue"
/lab_host="DH10B"
/note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo
Average insert 2.5 kb. Library constructed by Life
Technologies, catalog # 12018-016. Investigators providing
samples: Lothar Hennighausen/Priscilla Furth, NIH
Reference for transgenic model: Li et al., Cell Growth and
Differentiation 7, 3-11 (1996). Note: this is a NIH_CGAP
Library."
BASE COUNT 134 a 202 c 176 g 133 t
ORIGIN

Query Match 96.9%; Score 532.2; DB 11; Length 645;
Best Local Similarity 99.3%; Pred. No. 1.5e-130;
Matches 545; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

Qy 1 atgtttgttgcctgtgctctcctgtgcctcctggtgtggtcgcacactcctcagtttcac 60
Db 66 ATGTTGTCTGCTGTGCTCTCTTGGCTCTGCTGGTGTGGCAACTCTCTGAGTTTATC 125
Qy 61 gtgcctccagtgagtgagggcctgcctccatccagtgctctagccgctggggcaccca 120
Db 126 GTGCCCCGATGAGTGTGAGGGCCCTGTCATCCGAGTGTCTAGCCGCTGGGCAACCA 185
Qy 121 gttcgtcacgtggtgtatctcacacacagccgagctctctcgaacagccgagactctgt 180
Db 186 GTTCGCTACGTGGTGAATCTACACACACCCGGCAGCTTCTCAACAGCCCGGACTCTGT 245
Qy 181 gaacagcagggcccgcaatgtgcagcattaccacaagaatgagctgggctggtgcagtga 240

Db 246 GAACAGCAGCCCGCATGTGAGCATATACCAAGAAATAGCTGGCTGGTGGATGTA 305
QY 241 gctacaactctcttatttgagagagcggtcgtctatgataagagccgagctggaacatc 300
Db 306 GCCTACAACCTTCTTATTGAGAGAGCGGTCTATGTCTATGAAGCCGAGGCTGAACATC 365
QY 301 aaggtgtgacacacagggcccatctggaatcccatgtctatttgatccacttcacgtggg 360
Db 366 AAGGTTGACACACAGAGGCCCATCTGGAATCCCATGTCTATTGGCATCACTTTCATGGGG 425
QY 361 aactcatgagccgggtaccccgcaagcgggccctccgtcgtcgtcctcctcctcctcctc 420
Db 426 AACTTCATGACCGGTACCCGCAAAAGCGGCCCTCGGTGCTCCCTTAATCTTCTGGAA 485
QY 421 tttgggtgtctcgtgggtctcctcgtgagatccaatatgaagtcaaaagacacccggatgtg 480
Db 486 TGTGGGGTGTCTCTGGGGCTTCTCTGAGATCCACTATGAGTCAAGAGGACACCGGATGTG 545
QY 481 caaagcaactctctcctcaggtgacaaactctcaggtcatcctcctcctcctcctcctc 539
Db 546 CAAGCACTCTCTCTCCAGTGACCAACTCTATCAGGTATCCAAAGCTGGGAACACTA 605
QY 540 ccgagagtgtg 548
Db 606 CCGAGAGTG 614

RESULT 10
LOCUS BI409815 842 bp mRNA 14-AUG-2001
DEFINITION 602961906F1 NCI_CGAP_Lu33 Mus musculus cDNA clone IMAGE:5117688 5',
rna sequence.
ACCESSION BI409815
VERSION BI409815.1 GI:15170738
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 842)
TITLE NIH-MGC http://mgc.nci.nih.gov/.
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM11286 row: k column: 01
High quality sequence stop: 568.
Location/Qualifiers
1. 842
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:5117688"
/clone_lib="NCI_CGAP_Lu33"
/tissue_type="pooled lung tumors"
/lab_host="DH10B (phage-resistant)"
/note="Organ: lung; Vector: pT73D-Pac (Pharmacia) with a
modified polylinker; Site_1: NotI; Site_2: EcoRI; 1st
strand cDNA was prepared from mRNA obtained from pooled
lung tumors with a Not I - oligo(dT) primer [5',
TGTACCAATCTGAGTGGGAGCGGCTGCTGTTTCTTTTCTTTT 3'].
Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not
I and Eco RI sites of the modified pT73 vector. Library

```

BASE COUNT 201 a 259 c 226 g 154 t 2 others
 ORIGIN
 went through one round of normalization, and was constructed by Bento Soares and M. Fatima Bonaldo.

Query Match 96.9%; Score 532.2; DB 11; Length 842;
 Best Local Similarity 99.1%; Pred. No. 1.6e-130;
 Matches 534; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 11 cctgtactctccttgcctctggtgtgcaacctcctgcagtttcatcgtgccccga 70
 Db 1 CTTGTCTCTCTCTCGCCCTTATGGTGTGCAACCTCTCTGAGTTTCATCGTCCCGCA 60
 QY 71 gtgagtgagggccctgcctcagctgtctagcgcctgtggggaacccagttcgtacg 130
 Db 61 GTGAGTGGAGGGCCCTGCCATCGAGTGTCTAGCGCGCTGGGGCACCCAGTTCGCTAGG 120
 QY 131 tgggtatctcacacagccgagcgtcttctgcaacagccgagcactcctgtgaacagcagg 190
 Db 121 TGGTGTATCTCACACAGCGCGCAGCTTCTGCAACAGCCCGGACTCTCTGTGAACAGCAGG 180
 QY 191 cccgcaatgtgagcattacacagaagaatgagctgggtgtgctgtagctacaact 250
 Db 181 CCGCAATGTGAGCATTACCAACAAGATGAGTGGGTGTGCGATGTAGCTTCAACT 240
 QY 251 tctttattggagagcagctcgtctgtatgaagccgagcgtggaacatcaaggtagcc 310
 Db 241 TCTTTANTGAGAGGACGGTGTATGTAAGCCGAGGCTTGGAACTATCAAGGGTGAAC 300
 QY 311 acacagggcccatctggaatccatctctattggcattcaccttctatgggaacttcagg 370
 Db 301 ACACAGGGCCCATCTTGGAACTCCATGTCTANTGGCATCACCTTTCATGG 360
 QY 371 accgggtaccggcaagggccctccctgctgctccctaaatcttctggaatgtgggtgt 430
 Db 361 ACCGGGTACCCGCAAGGGGGCCCTCCGTCGCGCCCTAAATCTTCTGGAATGTGGGGGTG 420
 QY 431 ctctgggtctctctgagatcccaactatgaatgaagcaccgggagtgcaaacactc 490
 Db 421 CTCGGGGCTCTCTGAGATCCCAACTATGAAGTCAAGGACACCGGATGTGCAAGCAGCTC 480
 QY 491 tctctcaggtagcaactctatcaggtcattcccaagctgggaacactaccagagtgga 549
 Db 481 TCTCTCAGGTGACCACTCTATCAGGTATCCCAAGCTGGGAACACTTACCGAGAGTGA 539

RESULT 11
 BF302505 619 bp mRNA EST 21-NOV-2000
 LOCUS 602031534F1 NCI_CGAP_Sg2 Mus musculus cDNA clone IMAGE:4166360 5',
 DEFINITION mRNA sequence.
 ACCESSION BF302505
 VERSION BF302505.1 GI:11249045
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 619)
 NIH-MGC http://mgi.nci.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs@mail.nih.gov
 Tissue Procurement: Jeffrey E. Green, M.D.
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
 http://image.llnl.gov
 Plate: LLM9455 row: d column: 09

High quality sequence stop: 607.
 Location/Qualifiers
 1.619

/organism="Mus musculus"
 /strain="FVB/N"
 /db_xref="taxon:10090"
 /clone="IMAGE:4166360"
 /clone.lib="NCI_CGAP_Sg2"
 /lab_host="DH10B (T1 phage-resistant)"
 /note="Organ: salivary gland; Vector: pCMV-SPORT6; Site: 1;
 Not; Site: 2; Salt; Cloned unidirectionally. Primer: Oligo
 dt. Average insert size 1.3 kb. Constructed by Life
 Technologies. Note: this is a NCI_CGAP Library."
 BASE COUNT 132 a 185 c 169 g 133 t
 ORIGIN

Query Match 96.2%; Score 528.4; DB 11; Length 619;
 Best Local Similarity 99.6%; Pred. No. 1.5e-129;
 Matches 540; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 8 ttgctgtgtctcttgcctcctggtctggaacctcctgcagtttcatcgtgcccc 67
 Db 50 TTGCTGTGTCTCTCTGCTCTCTGCTGTGGCAACCTCTCTGCAAGTTCATCGTGGCCC 109
 QY 68 gcagtgtgtgagggccctgcctcctgagtgctctagccgctggggcaccagttcgt 127
 Db 110 GCAGTGTGTGAGGGCCCTGCGATCCGAGTGTCTAGCGCCTGGGGCACCAGTTCGCT 169
 QY 128 acgtgtgtatctcacacagccgagcgtcttctgcaacagcccgagcctcctgtgaacagc 187
 Db 170 ACGTGTGTATCTCACACAGCCGGCAGCTTCTGCAACAGCCCGGACTCTCTGTGAACAGC 229
 QY 188 agggccgcaatgtgcagcattaccacaagaatgagctgggtgtgctgagctatgacctaca 247
 Db 230 AGGCCGCAATGTGCAGATTACCAAGAATGAGCTGGGTGTGCGATGTAGCCTACA 289
 QY 248 attccttattggagagcagcgtcatgtctatgaagccgaggtgtgaacatcaagggtg 307
 Db 290 ACTTCTCTTA-TGGAGAGGACGGTCTATGTATGAAGCCGAGGCTGGAACATCAAGGGTG 348
 QY 308 accacagggccctatctggaatcccatgctctattgcatcaccctcctcattggaactca 367
 Db 349 ACCACAGGGCCCATCTGGATCCCATGCTATTTGCAATCACCCTTCATGGGAACCTCA 408
 QY 368 tggaccgggtaccgcgcaaaagcgggcccctcctgctgcctcctctctctggaatgtgggg 427
 Db 409 TGGACCGGGTACCCGCAAGCGGGCCCTCCGTGCTGCCCTAAATCTTCTGGAATGTGGG 468
 QY 428 tgtctcgggcttctcctgagatcccaactatgaagcacaagcaccgggagtgcaaaagca 487
 Db 469 TGCTCGGGGCTTCTTGAGATCCAACTATGAAGTCAAGGACACCCGGGATGTGCAAGCA 528
 QY 488 ctctctccaggtgaccactctctcaggtcattccaaagctgggaacactaccagagt 547
 Db 529 CTCTCTCCAGGTGACCAACTCTATCAGGTCTCCTCAAGCTGGGAACACTACCGAGGT 588
 QY 548 ga 549
 Db 589 GA 590

RESULT 12
 BE913706 1045 bp mRNA EST 29-SEP-2000
 LOCUS 601669059F1 NCI_CGAP_Mam1 Mus musculus cDNA clone IMAGE:3969014 5',
 DEFINITION mRNA sequence.
 ACCESSION BE913706
 VERSION BE913706.1 GI:10411594
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 17, 2001, 07:49:10 ; Search time 165.03 Seconds
(without alignments)
2852.036 Million cell updates/sec

Title: US-09-462-625-1

Perfect score: 549

Sequence: 1 atgttttgcctgtgtctt.....gggaacactaccgagagtga 549

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N_Geneseq_1101.*
1: /SID88/gcgdata/geneseq/geneseq/NA1980.DAT.*
2: /SID88/gcgdata/geneseq/geneseq/NA1981.DAT.*
3: /SID88/gcgdata/geneseq/geneseq/NA1982.DAT.*
4: /SID88/gcgdata/geneseq/geneseq/NA1983.DAT.*
5: /SID88/gcgdata/geneseq/geneseq/NA1984.DAT.*
6: /SID88/gcgdata/geneseq/geneseq/NA1985.DAT.*
7: /SID88/gcgdata/geneseq/geneseq/NA1986.DAT.*
8: /SID88/gcgdata/geneseq/geneseq/NA1987.DAT.*
9: /SID88/gcgdata/geneseq/geneseq/NA1988.DAT.*
10: /SID88/gcgdata/geneseq/geneseq/NA1989.DAT.*
11: /SID88/gcgdata/geneseq/geneseq/NA1990.DAT.*
12: /SID88/gcgdata/geneseq/geneseq/NA1991.DAT.*
13: /SID88/gcgdata/geneseq/geneseq/NA1992.DAT.*
14: /SID88/gcgdata/geneseq/geneseq/NA1993.DAT.*
15: /SID88/gcgdata/geneseq/geneseq/NA1994.DAT.*
16: /SID88/gcgdata/geneseq/geneseq/NA1995.DAT.*
17: /SID88/gcgdata/geneseq/geneseq/NA1996.DAT.*
18: /SID88/gcgdata/geneseq/geneseq/NA1997.DAT.*
19: /SID88/gcgdata/geneseq/geneseq/NA1998.DAT.*
20: /SID88/gcgdata/geneseq/geneseq/NA1999.DAT.*
21: /SID88/gcgdata/geneseq/geneseq/NA2000.DAT.*
22: /SID88/gcgdata/geneseq/geneseq/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	549	100.0	549	AA21819	Mouse tag7 clone c
2	523.4	95.3	677	AA21819	Murine granulocyte
3	316.6	57.7	697	AA21819	Human PRO1269 nucl
4	316.6	57.7	697	AA21819	Human PRO1269 (UNQ)
5	316.6	57.7	697	AA21819	DNA encoding prote
6	316.6	57.7	718	AA21819	Human tag7 clone c
7	316.6	57.7	726	AA21819	Human Htag7 secret
8	316.6	57.7	749	AA21819	Chondrosarcoma pep
9	284.2	51.8	688	AA21819	Bovine granulocyte
10	155.6	28.3	380	AA21819	Human 5' EST isola
11	126.8	23.1	1107	AA21819	Human full length

12	126.8	23.1	1125	22	AA21819	Human full length
13	107	19.5	1110	22	AA21819	Human full length
14	107	19.5	1128	22	AA21819	Human full length
15	105.8	19.3	1876	21	AA21819	Wound healing tiss
16	84.4	15.4	285	21	AA21819	Human secreted pro
17	78.4	14.3	1182	21	AA21819	Keratinocyte pepti
18	78.4	14.3	1191	21	AA21819	Human secreted pro
19	75.8	13.8	1089	21	AA21819	Human protein clon
20	75.8	13.8	1256	21	AA21819	Human secreted pro
21	73.4	13.4	457	21	AA21819	Human secreted pro
22	71	12.9	522	21	AA21819	Gene encoding pept
23	71	12.9	591	19	AA21819	Gene encoding pept
24	71	12.9	753	19	AA21819	Gene encoding the
25	56.2	10.2	1876	22	AA21819	Murine peptidoglyc
26	49.6	9.0	2049	22	AA21819	C glutamicum codin
27	49.6	9.0	309400	22	AA21819	C glutamicum codin
28	41	7.5	342	22	AA21819	Human truncated Zg
29	39.2	7.1	1200	22	AA21819	Human peptidoglyc
30	37.4	6.8	2297	22	AA21819	Human secreted pro
31	36.4	6.6	2445	22	AA21819	P. putida KT2440-a
32	35.8	6.5	2801	22	AA21819	Human cDNA encodin
33	35.8	6.5	2955	22	AA21819	Human polynucleoti
34	35.8	6.5	3028	22	AA21819	Human polynucleoti
35	35.8	6.5	3039	22	AA21819	Human polynucleoti
36	35.8	6.5	3039	22	AA21819	Human polynucleoti
37	35.4	6.4	1611	20	AA21819	Human polynucleoti
38	35.4	6.4	4936	22	AA21819	PHA synthase nucle
39	35.4	6.4	6436	20	AA21819	Alcaligenes latus
40	34.6	6.3	2608	16	AA21819	DNA containing PHA
41	34.6	6.3	6226	22	AA21819	Partial ALK gene.
42	33.8	6.2	1160	20	AA21819	Human ALK gene.
43	33.8	6.2	1195	20	AA21819	Enterococcus faeca
44	33.8	6.2	1195	20	AA21819	M. tuberculosis an
45	33.8	6.2	8519	21	AA21819	M. tuberculosis re
						Human fatty acid s

ALIGNMENTS

RESULT 1
ID: AAX21819 standard; cDNA; 549 BP.

XX AAX21819;

XX 18-MAY-1999 (first entry)

XX Mouse tag7 clone coding sequence.

XX Tag7; tumour growth inhibitor; mammalian tumour; carcinoma; sarcoma;

XX melanoma; leukaemia; apoptosis inducer; mouse; ds.

XX Mus sp.

XX WO9902686-A1.

XX 21-JAN-1999.

XX 10-JUL-1998; 98WO-EP04287.

XX 11-JUL-1997; 97US-0893764.

XX (BOEH) BOEHRINGER INGELHEIM INT GMBH.

XX Georgiev G, Kisslev S, Ostermann E, Prokhorchouk E;

XX WPI: 1999-120887/10.

XX P-FSDB; AAY00770.

XX New nucleic acid encoding tag7 - used to inhibit tumour growth and

XX induce apoptosis, for treatment of carcinoma, sarcoma, melanoma and

XX leukaemia

Claim 3; Fig 1; 138pp; English.

This sequence encodes the murine tag7 of the invention. Cells containing the tag7 DNA sequence are used to express recombinant tag7. Tag7 is used to produce and purify antibodies; to inhibit growth of mammalian tumours, especially for treating carcinoma (of liver, ovary, breast, cervix, lung, prostate, colon/rectum, bladder, testis, stomach, pancreas, mouth, head, and neck, squamous cell carcinoma or teratocarcinoma), sarcoma (Kaposi's, osteo- or fibro-sarcomas), melanoma or leukaemia; and as a molecular weight marker. The tag7 polypeptide inhibits tumour growth and induces apoptosis. The tag7 coding sequences are also useful as probes for gene mapping and detection of tag7 gene expression, and as primers. Antibodies against tag7 are used as reagents for detecting tag7; as an antagonist of tag7; for isolating tag7 and therapeutically to inhibit or delay tumour metastasis.

Sequence 549 BP; 117 A; 161 C; 151 G; 120 T; 0 other;

Query Match	100.0%;	Score 549;	DB 20;	Length 549;
Best Local Similarity	100.0%;	Pred. NO. 8.7e-156;		
Matches 549;	Conservative	0;	Mismatches 0;	Indels 0;
				Gaps 0;

[illegible]

RESULT 2
AAT78510
ID AAT78510 standard; cDNA; 677 BP.

AC AAT78510;

DT 18-FEB-1998 (first entry)

DE Murine granulocyte peptide A precursor cDNA.

XX	Antimicrobial peptide; antibiotic; antibacterial; antifungal;
KW	fungicide; antiprotozoa; protozoacide; antiviral; virucide;
KW	murine granulocyte peptide A; Mgp-A; preservative; sepsis;
KW	endotoxaemia; mouse; ss.
XX	
OS	Mus musculus.
XX	
PH	Key Location/Qualifiers
FT	CDS 36..581
FT	/*tag= a
FT	/transl_except= (pos:561..563, aa:Trp)
FT	mat_peptide 543..578
FT	/*tag= b
FT	/product= MGP-A
XX	
PN	WO9729765-A1.
XX	
PD	21-AUG-1997.
XX	
PD	13-FEB-1997; 97WO-US02218.
XX	
PF	16-FEB-1996; 96US-0011834.
XX	
PR	(REGC) UNIV CALIFORNIA.
XX	
PA	Selsted ME;
PI	WPI: 1997-424753/39.
XX	
DR	P-PSDB; AAW23723.
XX	
PT	Antimicrobial (poly)peptide(s) from bovine and murine granulocytes
PT	useful therapeutically, as preservatives for food, in water
PT	treatment and in agriculture
XX	
PS	Claim 14; Fig 5; 56pp; English.
XX	
CC	This cDNA sequence encodes the precursor (see AAW23723) of a novel,
CC	claimed antimicrobial peptide from bovine neutrophils, designated
CC	murine granulocyte peptide A or MGP-A (see AAW23725). It was
CC	isolated from murine bone marrow cDNA using primers based on
CC	bovine BGP-A cDNA (see AA78509). The encoded MGP-A precursor
CC	comprises signal peptide and propeptide sequences followed by 12
CC	C-terminal amino acids corresponding to mature MGP-A. MGP-A and
CC	BGP-A (see AAW23724), exhibit activity against Gram-positive and
CC	Gram-negative bacteria, fungi and viruses, specifically
CC	Staphylococcus aureus, Escherichia coli, Candida albicans,
CC	Salmonella typhimurium and C. neoformans (claimed). They can be
CC	used in human or veterinary medicine (particularly to treat
CC	disorders associated with lipopolysaccharides, e.g. sepsis and
CC	endotoxaemia) or as preservatives in food products or in water
CC	supplies (claimed). They can also be applied to crops to increase
CC	post-harvest spoilage or expressed in transgenic plants to increase
CC	their disease resistance. They have low immunogenicity.
XX	
SQ	Sequence 677 BP; 146 A; 217 C; 171 G; 143 T; 0 other;

Query Match	95.3%	Score 523.4;	DB 18;	Length 677;
Best Local Similarity	98.4%	Pred. No. 4.9e+148;		
Matches 540;	Conservative	0;	Mismatches 6;	Indels 3;
Gaps 1;				

Qy	1	atgtgttttgctgtgctctctctgtgctctgctgctgctgctgcaacctctgcagtttcac	60
Db	36	atgtgttttgctgtgctctctgtgctctgctgctgctgctgcaacctctgcagtttcac	95
Qy	61	gtgcccgcgagtgagtgaggggccctgccaatccgagtgctctctagccgcctggggcacc	120
Db	96	gtgcccgcgagtgagtgaggggccctgccaatccgagtgctctagccgcctggggcacc	155
Qy	121	gttcgcgtacgtgggtgattctcaacacagccggcagctctctgcaacagccggactcctgt	180
Db	156	gttcgcgtacgtgggtgattctcaacacagccggcagctctctgcaacagccggactcctgt	215

QY 181 gaacagcagccccaatgtgcagcattaccacagaatgagctgggctgtgcgatga 240
 Db 216 gaacagcagccccaatgtgcagcattaccacagaatgagctgggctgtgcgatga 275
 QY 241 gcctacaacttcctattggagagcagcgtctatgtatgaagccgagcgtggaacac 300
 Db 276 gcctacaacttcctattggagagcagcgtctatgtatgaagccgagcgtggaacac 335
 QY 301 aaggggtgaccacagagcccatctggaatcccatgtctattgcatcattcattggg 360
 Db 336 aaggggtgaccacagagcccatctggaatcccatgtctattgcatcattcattggg 395
 QY 361 aactctatgacccgggtaccgcgaagcggccctccgtgctgcctaaattcttggaa 420
 Db 396 aactctatgacccgggtaccgcgaagcggccctccgtgctgcctaaattcttggaa 452
 QY 421 tgggggtgtcgggcttctcctgagatcaactatgaagtcacacagccggatgtg 480
 Db 453 tcgggggtgtcgggcttctcctgagatcaactatgaagtcacacagccggatgtg 512
 QY 481 caaagcaactctctccaggtgacaaactctatcaggtcattcccaagcgtggaacactac 540
 Db 513 caaagcaactctctccaggtgacaaactctatcaggtcattcccaagcgtggaacactac 572
 QY 541 cgagagtga 549
 Db 573 cgagagtga 581

RESULT 3
 AAC58104
 ID AAC58104 standard; cDNA; 697 BP.
 AC AAC58104;
 AC AAC58104;
 DT 25-JAN-2001 (first entry)
 DE Human PRO1269 nucleotide sequence SEQ ID NO:6.
 DE Human: tumour; diagnosis; neoplastic disease; proliferation; cancer;
 KW identification; tumourigenesis; anticancer; detection; ss.
 KW Homo sapiens.
 OS WO200053750-A1.
 PN 14-SEP-2000.
 PD 02-DEC-1999; 99WO-US28551.
 PF 08-MAR-1999; 99WO-US05028.
 PR 01-SEP-1999; 99WO-US20111.
 PR 29-OCT-1999; 99US-0162506.
 PR 30-NOV-1999; 99WO-US28313.
 PR 01-DEC-1999; 99WO-US28634.
 XX (GETH) GENENTECH INC.
 PI Botstein D, Goddard A, Gurney AL, Roy MA, Watanabe CK, Wood WI;
 DR WPI; 2000-594320/56.
 DR P-PSDB; AAB24022.
 XX Antibodies specific for PRO polypeptides, used to diagnose and inhibit
 PT the growth of tumors in mammals, and to identify inhibitors of PRO
 PT polypeptide activity or expression -
 XX Claim 50; Fig 3; 226pp; English.
 XX The present invention describes an antibody that binds to a human
 CC protein (I) selected from: PRO381; PRO1269; PRO1410; PRO1755; PRO1780;
 CC PRO3434; PRO1927; PRO3567; PRO1295; PRO1293; PRO1303; PRO344; PRO4354;

CC PRO4397; PRO4407; PRO1555; PRO1096; PRO2038; and PRO2262. (I) has
 CC anticancer activity and can be used to diagnose tumours in mammals, by
 CC detecting complex formation when the antibody is contacted with test
 CC cells. Increased expression of genes encoding (I) can also be detected
 CC to diagnose tumours. Agents which inhibit the activity of (I),
 CC especially the antibodies, or an antisense oligonucleotide which
 CC hybridises to genes encoding (I), can be used to inhibit tumour growth,
 CC preferably by inducing cell death. Methods from the present invention
 CC can be used to identify compounds which inhibit the biological activity
 CC of (I). AAC58019 to AAC58102 represent PCR primers and hybridisation
 CC probes used in examples from the present invention for human PRO
 CC sequences. AAC58103 to AAC58122 and AAB24021 to AAB24040 represent human
 CC PRO polynucleotide and protein sequences given in the exemplification of
 CC the present invention.
 XX
 SQ Sequence 697 BP; 136 A; 250 C; 184 G; 127 T; 0 other;
 Query Match 57.7%; Score 316.6; DB 21; Length 697;
 Best Local Similarity 77.2%; Pred. No. 1.1e-85;
 Matches 385; Conservative 0; Mismatches 114; Indels 0; Gaps 0;
 QY 44 cctcctgagtttcacgtgcccgcagtgagtgaggcctccatccagtgctcta 103
 Db 108 cctgctgagcccatagtgcccggagcagtgaggccctgcacagatgagtgcc 167
 QY 104 gccgcctggggcaacccagttcgtacgtggtggtatctcacacagccggcagcttctga 163
 Db 168 agcactcgtgctgcccctacgtatgtggtggtatcgcacacgctggcagcagctga 227
 QY 164 acgcccgcactcctgtgaacagcagcggcccaatgtgcagcattaccacagaatgagc 223
 Db 228 acaccccgctcgtgccagcagcagcggcccgaaatgtgcagcactcaccacatgaagac 287
 QY 224 tgggctggtgcgtatgagcctacacttctctatttggagagcaggtcatgtctatgaag 283
 Db 288 tgggctggtgcagcgtgggtcacacttctctgattggagagcggctcgtatacagag 347
 QY 284 gccagcgtggaacatacaagggtgacacacagcggcccatctggaatcccatgtctattg 343
 Db 348 gccgtggtggaacttcacgggtgcccactcaggtcacttatggaacccatgtccattg 407
 QY 344 gcataccttcattgggaacttcacgggtgacccgggtacccgcaagcggccctccgtctg 403
 Db 408 gcaacagcttcattgggaacttcacgggtgacccgggtgcccacaccccgagccatccggcag 467
 QY 404 ccctaaatcttctggaatgtgggtgtctcgggtgtctcctgagatcccaactatgaagta 463
 Db 468 ccaggtgtactggtcgtggtgtggtcagagcggcctgaggtcccaactatgtctca 527
 QY 464 aaggacacccgggtgtgcaaaagcactctctctccaggtgacaaactctatcaggtctatcc 523
 Db 528 aaggacacccgggtgtgcaagcgttacactctctccaggtgacaaactctatcaggtctatcc 587
 QY 524 aagcgtggaacactacgg 542
 Db 588 agaattggccacactaccg 606

RESULT 4
 AAA37082
 ID AAA37082 standard; cDNA; 697 BP.
 XX
 AC AAA37082;
 XX
 DT 08-AUG-2000 (first entry)
 XX
 DE Human PRO1269 (UNQ639) cDNA sequence SEQ ID NO:215.
 XX
 KW Human; PRO polypeptide; membrane bound protein; receptor; diagnosis;
 KW transmembrane; secretion; immunoadhesion; pharmaceutical; screening;
 KW ss.
 XX

```
OS Homo sapiens.
XX WO200012708-A2.
XX PD 09-MAR-2000.
XX PF 01-SEP-1999; 99WO-US20111.
XX PR 01-SEP-1998; 98US-0098716.
PR 01-SEP-1998; 98US-0098749.
PR 01-SEP-1998; 98US-0098750.
PR 02-SEP-1998; 98US-0098803.
PR 02-SEP-1998; 98US-0098821.
PR 02-SEP-1998; 98US-0098843.
PR 09-SEP-1998; 98US-0099536.
PR 09-SEP-1998; 98US-0099596.
PR 09-SEP-1998; 98US-0099598.
PR 09-SEP-1998; 98US-0099602.
PR 09-SEP-1998; 98US-0099642.
PR 10-SEP-1998; 98US-0099741.
PR 10-SEP-1998; 98US-0099754.
PR 10-SEP-1998; 98US-0099763.
PR 10-SEP-1998; 98US-0099792.
PR 10-SEP-1998; 98US-0099808.
PR 10-SEP-1998; 98US-0099812.
PR 10-SEP-1998; 98US-0099815.
PR 10-SEP-1998; 98US-0099816.
PR 15-SEP-1998; 98US-0100385.
PR 15-SEP-1998; 98US-0100388.
PR 15-SEP-1998; 98US-0100390.
PR 16-SEP-1998; 98US-0100584.
PR 16-SEP-1998; 98US-0100627.
PR 16-SEP-1998; 98US-0100661.
PR 16-SEP-1998; 98US-0100662.
PR 16-SEP-1998; 98US-0100664.
PR 17-SEP-1998; 98US-0100683.
PR 17-SEP-1998; 98US-0100684.
PR 17-SEP-1998; 98US-0100710.
PR 17-SEP-1998; 98US-0100711.
PR 17-SEP-1998; 98US-0100919.
PR 17-SEP-1998; 98US-0100930.
PR 18-SEP-1998; 98US-0100848.
PR 18-SEP-1998; 98US-0100849.
PR 18-SEP-1998; 98US-0101014.
PR 18-SEP-1998; 98US-0101068.
PR 18-SEP-1998; 98US-0101071.
PR 22-SEP-1998; 98US-0101279.
PR 23-SEP-1998; 98US-0101471.
PR 23-SEP-1998; 98US-0101472.
PR 23-SEP-1998; 98US-0101474.
PR 23-SEP-1998; 98US-0101475.
PR 23-SEP-1998; 98US-0101476.
PR 23-SEP-1998; 98US-0101477.
PR 23-SEP-1998; 98US-0101479.
PR 24-SEP-1998; 98US-0101738.
PR 24-SEP-1998; 98US-0101741.
PR 24-SEP-1998; 98US-0101743.
PR 24-SEP-1998; 98US-0101915.
PR 24-SEP-1998; 98US-0101916.
PR 29-SEP-1998; 98US-0102207.
PR 29-SEP-1998; 98US-0102240.
PR 29-SEP-1998; 98US-0102307.
PR 29-SEP-1998; 98US-0102330.
PR 29-SEP-1998; 98US-0102331.
PR 30-SEP-1998; 98US-0102484.
PR 30-SEP-1998; 98US-0102487.
PR 30-SEP-1998; 98US-0102570.
PR 30-SEP-1998; 98US-0102571.
PR 01-OCT-1998; 98US-0102684.
PR 01-OCT-1998; 98US-0102687.
PR 02-OCT-1998; 98US-0102965.
PR 06-OCT-1998; 98US-0103258.
PR 06-OCT-1998; 98US-0103449.
PR 07-OCT-1998; 98US-0103314.
PR 07-OCT-1998; 98US-0103315.
PR 07-OCT-1998; 98US-0103328.
PR 07-OCT-1998; 98US-0103395.
PR 07-OCT-1998; 98US-0103396.
PR 07-OCT-1998; 98US-0103401.
PR 08-OCT-1998; 98US-0103633.
PR 08-OCT-1998; 98US-0103678.
PR 08-OCT-1998; 98US-0103679.
PR 14-OCT-1998; 98US-0103711.
PR 14-OCT-1998; 98US-0104257.
PR 20-OCT-1998; 98US-0104987.
PR 20-OCT-1998; 98US-0105000.
PR 20-OCT-1998; 98US-0105002.
PR 21-OCT-1998; 98US-0105104.
PR 22-OCT-1998; 98US-0105169.
PR 22-OCT-1998; 98US-0105266.
PR 26-OCT-1998; 98US-0105693.
PR 26-OCT-1998; 98US-0105694.
PR 27-OCT-1998; 98US-0105807.
PR 27-OCT-1998; 98US-0105881.
PR 27-OCT-1998; 98US-0105882.
PR 27-OCT-1998; 98US-0106062.
PR 28-OCT-1998; 98US-0106023.
PR 28-OCT-1998; 98US-0106029.
PR 28-OCT-1998; 98US-0106030.
PR 28-OCT-1998; 98US-0106032.
PR 28-OCT-1998; 98US-0106033.
PR 28-OCT-1998; 98US-0106178.
PR 29-OCT-1998; 98US-0106248.
PR 29-OCT-1998; 98US-0106384.
PR 29-OCT-1998; 98US-0108500.
PR 30-OCT-1998; 98US-0106464.
PR 03-NOV-1998; 98US-0106856.
PR 03-NOV-1998; 98US-0106902.
PR 03-NOV-1998; 98US-0106905.
PR 03-NOV-1998; 98US-0106919.
PR 03-NOV-1998; 98US-0106932.
PR 03-NOV-1998; 98US-0106934.
PR 10-NOV-1998; 98US-0107783.
PR 17-NOV-1998; 98US-0108775.
PR 17-NOV-1998; 98US-0108779.
PR 17-NOV-1998; 98US-0108787.
PR 17-NOV-1998; 98US-0108788.
PR 17-NOV-1998; 98US-0108801.
PR 17-NOV-1998; 98US-0108802.
PR 17-NOV-1998; 98US-0108806.
PR 17-NOV-1998; 98US-0108807.
PR 17-NOV-1998; 98US-0108867.
PR 18-NOV-1998; 98US-0108925.
PR 18-NOV-1998; 98US-0108849.
PR 18-NOV-1998; 98US-0108850.
PR 18-NOV-1998; 98US-0108851.
PR 18-NOV-1998; 98US-0108852.
PR 18-NOV-1998; 98US-0108858.
PR 18-NOV-1998; 98US-0108904.
XX (GETH ) GENENTECH INC.
XX Baker K, Goddard A, Gurney AL, Smith V, Watanabe CK, Wood WI;
XX WPI; 2000-237871/20.
XX P-PSDB; AAY99400.
XX New mammalian DNA sequences encoding transmembrane, receptor or
PT secreted PRO polypeptides, useful for screening of potential peptide or
PT small molecule inhibitors of the relevant receptor/ligand interactions
XX Claim 2; Fig 121; 773pp; English.
XX AAA37022 to AAA37144 encode the new isolated human transmembrane,
CC receptor or secreted PRO polypeptides given in AAY99340 to AAY99462. The
```

transmembrane and receptor PRO proteins can be used for screening of potential peptide or small molecule inhibitors of the relevant receptor/ligand interactions. The polypeptides and nucleotide sequences encoding them have various industrial applications, including uses as pharmaceutical and diagnostic agents. AAA37145 to AAA37330 represent PCR primers and hybridization probes used in the isolation of the PRO polypeptides from the present invention.

Sequence 697 BP; 136 A; 250 C; 184 G; 127 T; 0 other;

Query Match 57.7%; Score 316.6; DB 21; Length 697;
Best Local Similarity 77.2%; Pred. No. 1.1e-85;
Matches 385; Conservative 0; Mismatches 114; Indels 0; Gaps 0;

44 cctcctgagtttcctgcccgcagtgagtgaggccctgcattccagtgctcta 103
108 cctgctgagcccatagtcgccggaacagtggaagccctggatcagtgagcgc 167
104 gccgctggggcaccagttcgtacgtggtgatctcacacagccgagcttctga 163
168 agcaccctgagctcccttgcattggtggtgatcgacacgcgggcagcagctga 227
164 acgcccggactcctgtgaacagcagggcccgcaatgtgcagcattaccacaagaatgagc 223
228 acaccccgctcgtgccagcagcagggcccggaatgtgcagcattaccacaagaatgagc 287
224 tgggctggtgcagtgatgagctcattcttattgagagcagtgctatgaag 283
288 tgggctggtgcagtgatgagctcattcttattgagagcagtgctatgaag 347
284 gccagctgggaacatacaagggtgacacacagggcccatctggaatcccatgtctattg 343
348 gccgtggctggaactcaccgggtgcccactcagtgctattggaaccccatgtccattg 407
344 gcattcactctcattgggaactcattgacccgggtaccgcaagcgggcccctcgtgctg 403
408 gcattcactctcattgggaactcattgacccgggtaccgcaagcgggcccctcgtgctg 467
404 ccttaattctctggaatgtggggtgctcgggcttctcagagatccaaactatgaatga 463
468 cccaggtctactgacctgctgctgctcagggagccctgaggtccaaactatgtgctca 527
464 aaggacacgggagtgcgaagacactctctcaggtgacacaaactatcagggtcatcc 523
528 aaggacacgggagtgcgaagacactctctcaggtgacacaaactatcagggtcatcc 587
524 aagctgggaacactaccg 542
588 agaattggccacactaccg 606

RESULT 5

AAF54356
ID AAF54356 standard; DNA; 697 BP.

AC AAF54356;

XX 02-APR-2001 (first entry)

DE DNA encoding protein of the invention #61.

DE Secreted; transmembrane; gene therapy; ss.

XX Unidentified.

XX WO200078961-A1.

XX 28-DEC-2000.

XX 18-FEB-2000; 2000WO-US04342.

XX 23-JUN-1999; 99US-0141037.

20-JUL-1999; 99US-0144758.
26-JUL-1999; 99US-0145698.
01-SEP-1999; 99WO-US20111.
29-OCT-1999; 99US-0162506.
30-NOV-1999; 99WO-US28313.
02-DEC-1999; 99WO-US28551.
16-DEC-1999; 99WO-US30095.
05-JAN-2000; 2000WO-US00219.
06-JAN-2000; 2000WO-US00376.
XX (GETH) GENENTECH INC.
PA Baker KP, Botstein D, Desnoyers L, Eaton DL, Ferrara N, Fong S;
XX Gao W, Goddard A, Godowski PJ, Grimaldi CJ, Gurney AL, Hillan KJ;
PI Pan J, Paoni NF, Roy MA, Smith V, Stewart TA, Tumas D;
PI Watanabe CK, Williams PM, Wood WI;
XX WPI; 2001-071395/08.

Secreted and transmembrane proteins and nucleic acids designated PRO, useful as hybridization probes, in chromosome and gene mapping and gene therapy -

Claim 2; Fig 121; 787pp; English.

The present invention relates to secreted and transmembrane proteins. These proteins and the DNA encoding them may be used as hybridization probes, in chromosome and gene mapping and in the generation of anti-sense RNA and DNA. They may also be used to generate either transgenic animals or knockout animals which are in turn useful for development and screening of therapeutically useful reagents. The nucleic acids may also be used in gene therapy.

Sequence 697 BP; 136 A; 250 C; 184 G; 127 T; 0 other;

Query Match 57.7%; Score 316.6; DB 22; Length 697;
Best Local Similarity 77.2%; Pred. No. 1.1e-85;
Matches 385; Conservative 0; Mismatches 114; Indels 0; Gaps 0;

44 cctcctgagtttcctgcccgcagtgagtgaggccctgcattccagtgctcta 103
108 cctgctgagcccatagtcgccggaacagtggaagccctggatcagtgagcgc 167
104 gccgctggggcaccagttcgtacgtggtgatctcacacagccgagcttctga 163
168 agcaccctgagctcccttgcattggtggtgatcgacacgcgggcagcagctga 227
164 acgcccggactcctgtgaacagcagggcccgcaatgtgcagcattaccacaagaatgagc 223
228 acaccccgctcgtgccagcagcagggcccggaatgtgcagcattaccacaagaatgagc 287
224 tgggctggtgcagtgatgagctcattcttattgagagcagtgctatgaag 283
288 tgggctggtgcagtgatgagctcattcttattgagagcagtgctatgaag 347
284 gccagctgggaacatacaagggtgacacacagggcccatctggaatcccatgtctattg 343
348 gccgtggctggaactcaccgggtgcccactcagtgctattggaaccccatgtccattg 407
344 gcattcactctcattgggaactcattgacccgggtaccgcaagcgggcccctcgtgctg 403
408 gcattcactctcattgggaactcattgacccgggtaccgcaagcgggcccctcgtgctg 467
404 ccttaattctctggaatgtggggtgctcgggcttctcagagatccaaactatgaatga 463
468 cccaggtctactgacctgctgctgctcagggagccctgaggtccaaactatgtgctca 527
464 aaggacacgggagtgcgaagacactctctcaggtgacacaaactatcagggtcatcc 523
528 aaggacacgggagtgcgaagacactctctcaggtgacacaaactatcagggtcatcc 587
524 aagctgggaacactaccg 542

Db 588 agaatggccacactaccg 606
 || ||| |||||

RESULT 6

AAX21820
 ID AAX21820 standard; cDNATO mRNA; 718 BP.

XX AAX21820;
 AC

XX 18-MAY-1999 (first entry)
 DT

XX Human tag7 clone coding sequence.
 DE

XX Tag7; tumour growth inhibitor; mammalian tumour; carcinoma; sarcoma;
 KW melanoma; leukaemia; apoptosis inducer; human; ss.

XX Homo sapiens.
 OS

XX WO9902686-A1.
 PN

XX 21-JAN-1999.
 PD

XX 10-JUL-1998; 98WO-EP04287.
 PF

XX 11-JUL-1997; 97US-0893764.
 PR

XX (BOEH) BOEHRINGER INGELHEIM INT GMBH.
 PA

XX Georgiev G, Kiselev S, Ostermann E, Prokhorchouk E;
 PI

XX WPI; 1999-120887/10.
 DR

XX P-PSDB; AAY00771.
 DR

XX New nucleic acid encoding tag7 - used to inhibit tumour growth and
 PT induce apoptosis, for treatment of carcinoma, sarcoma, melanoma and
 PT leukaemia

XX Claim 11: Page 126-127; 138pp; English.
 PS

XX This sequence encodes the human tag7 of the invention. Cells containing
 CC the tag7 DNA sequence are used to express recombinant tag7. Tag7 is used
 CC to produce and purify antibodies: to inhibit growth of mammalian tumours,
 CC especially for treating carcinoma (of liver, ovary, breast, cervix, lung,
 CC prostate, colon/rectum, bladder, testis, stomach, pancreas, mouth, head
 CC and neck, squamous cell carcinoma or teratocarcinoma), sarcoma (Kaposi's,
 CC osteo- or fibro-sarcomas), melanoma or leukaemia; and as a molecular
 CC weight marker. The tag7 polypeptide inhibits tumour growth and induces
 CC apoptosis. The tag7 coding sequences are also useful as probes for gene
 CC mapping and detection of tag7 gene expression, and as primers. Antibodies
 CC against tag7 are used as reagents for detecting tag7; as an antagonist of
 CC tag7; for isolating tag7 and therapeutically to inhibit or delay tumour
 CC metastasis.

XX Sequence 718 BP; 142 A; 251 C; 194 G; 131 T; 0 other;
 SQ

Query Match 57.7%; Score 316.6; DB 20; Length 718;
 Best Local Similarity 77.2%; Pred. No. 1.1e-85;
 Matches 385; Conservative 0; Mismatches 114; Indels 0; Gaps 0;

QY 44 cctctcagtttcatcgtgccccgcgagtgagtgaggccctccatccagtgctcta 103
 ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||

Db 135 cctgtcagccccatagtgcccggaacgagtgagtgaggccctccatccagtgctcta 194
 ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||

QY 104 gcccctggggcaccacagttcgtacgtggtgatctcacacacgcccgcagcttgca 163
 ||| ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 195 agcactgagctgcccctacgtatgtggtgatcgacacgcccgcgagcagctgca 254
 ||| ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 164 acagcccgagctctgtgaacagcagcccgagtgagtgagtgagtgagtgagtgagc 223
 ||| ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 255 acaccccgctctgcccagcagcagcccgagtgagtgagtgagtgagtgagtgagc 314
 ||| ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 224 tgggctggtgctgctagctacacaaacttcctattgagagggagcggctctctatgaag 283
 ||| ||||| ||||| ||| ||||| ||||| ||||| ||||| ||||| ||| ||||| |||
 Db 315 tgggctggtgctgctagctacacaaacttcctattgagagggagcggctctctatgaag 374
 ||| ||||| ||||| ||| ||||| ||||| ||||| ||||| ||||| ||| ||||| |||
 QY 284 gccgaggtggaacatcaagggtgaccacacagggcccatctggaatcccatgtctattg 343
 ||||| ||||| ||||| ||| ||||| ||||| ||||| ||||| ||||| ||| ||||| |||
 Db 375 gccgtggtggaacttcacgggtgcccactcaggtcacttatggaaccccatgtccattg 434
 ||||| ||||| ||||| ||| ||||| ||||| ||||| ||||| ||||| ||| ||||| |||
 QY 344 gcataccttcattggggaacttcacgggtgaccgggtgaccggcgaagcgggcccctcgtgctg 403
 ||||| ||||| ||||| ||| ||||| ||||| ||||| ||||| ||||| ||| ||||| |||
 Db 435 gcataccttcattggggaacttcacgggtgaccgggtgaccggcgaagcgggcccctcgtgctg 494
 ||||| ||||| ||||| ||| ||||| ||||| ||||| ||||| ||||| ||| ||||| |||
 QY 404 cctataatcttctggaatgtgggtgtctcgggtctcctgagatcccaactatgaagtcga 463
 ||| ||||| ||||| ||| ||||| ||||| ||||| ||||| ||||| ||| ||||| |||
 Db 495 cccaggtgtctactggcctgctggtggtgagcgggtgaggtcccaactatgtgctca 554
 ||||| ||||| ||||| ||| ||||| ||||| ||||| ||||| ||||| ||| ||||| |||
 QY 464 aaggacacgggtgtgcaaaagcactctctccaggtgaccacactctcaggtcacc 523
 ||||| ||||| ||||| ||| ||||| ||||| ||||| ||||| ||||| ||| ||||| |||
 Db 555 aaggacacgggtgtgcaaaagcactctctccaggtgaccacactctcaggtcacc 614
 ||||| ||||| ||||| ||| ||||| ||||| ||||| ||||| ||||| ||| ||||| |||
 QY 524 aaagctggggaacactaccg 542
 || ||| ||||| |||||
 Db 615 agaattggccacactaccg 633
 || ||| ||||| |||||

RESULT 7

AAA80613

ID AAA80613 standard; cDNA; 726 BP.

XX AAA80613;
 AC

XX 21-NOV-2000 (first entry)
 DT

XX Human Htag7 secreted protein gene #8.
 DE

XX Secreted protein; immunosuppressant; anti-inflammatory; antiarthritic;
 KW antirheumatic; dermatological; antiproliferative; antiarteriosclerotic;
 KW anticancer; vulnary; antiviral; antibacterial; antifungal;
 KW Immune disorder; Addison's disease; rheumatoid arthritis; dermatitis;
 KW multiple sclerosis; inflammatory disorder; inflammatory bowel disease;
 KW Crohn's disease; nephritis; hyperproliferative disorder;
 KW cardiovascular disorder; coronary arteriosclerosis; myocarditis; cancer;
 KW melanoma; lymphoma; wound healing; human; ss.

XX Homo sapiens.

XX WO200029435-A1.
 PN

XX 25-MAY-2000.
 PD

XX 27-OCT-1999; 99WO-US25031.
 PF

XX 28-OCT-1998; 98US-0105971.
 PR

XX (HUMA-) HUMAN GENOME SCI INC.
 PA

XX Ni J, Ruben SM, Olsen HS, Young PE, Kenny JJ, Moore PA, Wei Y;
 PI Greene JM;
 PI

XX WPI; 2000-387742/33.
 DR

XX P-PSDB; AAB25583.
 DR

XX Isolated nucleic acid molecules encoding human secreted proteins are
 PT used for the prevention, amelioration and treatment of autoimmune,
 PT inflammatory, hyperproliferative and cardiovascular disorders, cancer,
 PT wounds, and infectious diseases -
 XX

XX Claim 1; Figure 34; 803pp; English.
 PS

XX The present invention relates to 12 secreted human proteins and the
 CC nucleotide sequences encoding them. The polynucleotide sequences given
 CC in AAA80613-A80623 encode the 12 secreted protein sequences given in
 CC AAB25576-B25593. The human secreted proteins have various activities

CC dependent on the tissues in which they are expressed. Examples of the
 CC activities of the proteins include: immunosuppressant;
 CC anti-inflammatory; antiarthritic; antirheumatic, dermatological;
 CC anti-proliferative; antiarteriosclerotic; anticancer; vulnerary;
 CC antiviral; antibacterial; and antifungal activity. The proteins,
 CC polypeptides, agonists and antagonists may be used to treat prevent
 CC and/or diagnose various disease, disorders and conditions examples of
 CC which include: immune disorders e.g. Addison's disease, rheumatoid
 CC arthritis, dermatitis, and multiple sclerosis; inflammatory disorders
 CC e.g. inflammatory bowel disease, Crohn's disease and nephritis;
 CC hyperproliferative disorders such as paraproteinemias and purpura;
 CC cardiovascular disorders e.g. coronary arteriosclerosis and myocarditis;
 CC cancer e.g. melanoma and lymphoma. The proteins and polynucleotide
 CC sequences may also be used in wound healing and the treatment of
 CC infectious diseases. The human secreted protein gene #8 and protein
 CC sequences are represented in sequences AAA80613 and AAB25583. Sequences
 CC AAA80662-A80663 represent genes related to the secreted protein gene#8.
 XX
 SQ Sequence 726 BP; 161 A; 251 C; 188 G; 126 T; 0 other;

Query Match 57.7%; Score 316.6; DB 21; Length 726;
 Best Local Similarity 77.2%; Pred. No. 1.1e-85;
 Matches 385; Conservative 0; Mismatches 114; Indels 0; Gaps 0;

QY 44 cctctgcagttctcatcgtgccccgcagtgagtgaggccctgccatccgagtgctcta 103
 DB 114 cctgctgcagccccatagtgccccgaagagtggaagccctggcatcagagtgcccc 173
 QY 104 gccgctggggcaccagtgctgaagtggtgatctcacacagccggcagcttctga 163
 DB 174 agcactgagctgcctcactgctatgtgtgtgtatcgcacagcgggcagcagctga 233
 QY 164 acagcccgactcctgtgaacagcagccgcgaatgtgcagcattaccacaagaatgagc 223
 DB 234 acaccccgctcgtgcagcagcagccggaggaatgtgcagcactaccacatgaagacac 293
 QY 224 tgggctgtgcagtagctacacactctctattggagaggaagcagtgctctatgaag 283
 DB 294 tgggctgtgcagtggtggtcacaactctctgattggagaagcagggctcgtatcagag 353
 QY 284 gccgagcttggaacatcaggtgacacacagggcccatctggaaatcccatgtctatcg 343
 DB 354 gccgtggctggaaactcagcgggtgcccactcaggtcacttatggaaacccatgtccattg 413
 QY 344 gcatcactctcagtggaactcattgacccgggttacccgcgaagcggccctccgtgctg 403
 DB 414 gcatcagcttcagtggaactacatgagtcgggtgcccacacccacagccatcccgggcag 473
 QY 404 cctcaaatctctggaatgtgggtgtctcgggttctcagatcccaactatgaagtca 463
 DB 474 ccaggtctactgcgtgcgtgtggtcagggagccctgaggtcccaactatgtgctca 533
 QY 464 aaggacacgggatgtgcaagacactctctccaggtgacacactctatcaggtccatcc 523
 DB 534 aaggacacgggatgtgcaggtgacactctctccaggtgacacactctatcaggtccatcc 593
 QY 524 aaagctgggaacactaccg 542
 DB 594 agaattggccacactaccg 612

RESULT 8
 ID AAA51719
 XX AAA51719 standard; cDNA; 749 BP.
 AC
 XX AAA51719;
 DT 31-OCT-2000 (first entry)
 XX
 DE Chondrosarcoma peptidoglycan recognition protein-like protein cDNA.
 XX
 KW Peptidoglycan recognition protein-like protein; PGRP; PGRP-C; regulator;

KW chondrosarcoma; immunosuppressive; antibacterial; anti-apoptotic;
 KW tumour necrosis factor; TNF; anti-infectious; mononuclear phagocyte;
 KW inhibitor; protein co-ordinate data; ss.
 XX Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT CDS 55..645 /tag= a
 FT /product= PGRP-like protein
 FT sig_peptide 55..117 /tag= b
 FT mat_peptide 118..642 /tag= c
 XX
 XX WO200039327-A1.
 XX
 PD 06-JUL-2000.
 XX
 PF 22-DEC-1999; 99WO-US30736.
 XX
 PR 23-DEC-1998; 98US-0113809.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 XX Rosen CA, Ruben SM, Young PE, Olsen HS;
 XX WPI; 2000-452414/39.
 DR P-PSDB; AAY96964.
 XX
 XX Polynucleotide encoding peptidoglycan recognition protein-like protein,
 PT antibodies specific to it useful for preventing, treating conditions
 PT e.g. endotoxic shock and auto-immune disorders and infections in mammal
 XX
 PS Claim 1; Fig 3; 19lpp; English.
 CC
 CC Novel human peptidoglycan recognition protein-like proteins (PGRP)
 CC expressed by keratinocytes (PGRP-K), wound healing tissues (PGRP-W)
 CC or chondrosarcoma tissue (PGRP-C) are disclosed. The PGRP DNA and
 CC proteins are useful for preventing, treating or ameliorating a medical
 CC condition in a mammal (claimed). PGRP is useful in augmenting the immune
 CC system in such areas as immune recognition, antigen presentation and
 CC immune system activation. Antibodies or antagonists directed against
 CC these proteins may be useful in reducing or eliminating disorders
 CC associated with tumour necrosis factor (TNF) and TNF-like cytokines,
 CC such as endotoxic shock and autoimmune disorders and for treating
 CC infectious diseases including silicosis, sarcoidosis and idiopathic
 CC pulmonary fibrosis.
 XX
 SQ Sequence 749 BP; 165 A; 259 C; 195 G; 130 T; 0 other;

Query Match 57.7%; Score 316.6; DB 21; Length 749;
 Best Local Similarity 77.2%; Pred. No. 1.1e-85;
 Matches 385; Conservative 0; Mismatches 114; Indels 0; Gaps 0;

QY 44 cctctgcagttctcatcgtgccccgcagtgagtgaggccctgccatccgagtgctcta 103
 DB 137 cctgctgcagccccatagtgccccgaagcagtggaagccctggcatcagagtgcccc 196
 QY 104 gccgctggggcaccacgttcctcgtgtgatctcacacacagccggcagcttctga 163
 DB 197 agcactgagcctgcctcactcgtatgtgtgtatcgcacacgcccggcagcagctgca 256
 QY 164 acagcccgactcctgtgaacagcagggcccgcaatgtgcagcattaccacaagaatgagc 223
 DB 257 acaccccgctcgtgcccagcagcagggcccggaatgtgcagcactaccacatgaagacac 316
 QY 224 tgggctgtgagatagcctacaaactcctctattggagagcaggtcatgtctatgaag 283
 DB 317 tgggctgtgagcgtgggctacaaactcctctgattggagaagcagggctcgtatcagag 376
 QY 284 gccgaggtggaacatcaagggtgaccacacagggcccatctcggaaatcccatctctatg 343

[illegible]

RESULT 9

AAT78509
ID AAT78509 standard; cDNA; 688 BP.

XX
AC AAT78509:

XX
DT 18-FEB-1998 (first entry)

XX DE Bovine granulocyte peptide A pre

Antimicrobial peptide; antibiotics
fungal; antiprotozoal; protozoal
bovine granulocyte peptide A; BG
endotoxaemia; cattle; ss.

XX
OS Bos taurus.

XX	Key	Location/Qualifi
FH		

```
FT CDS
FT 29..601
/*tag= a
```

FT	sig_peptide	29..91
FT		/*tag= b

```
FT mat_peptide 560..598
FT /*tag= c
```

FT
XX
/product= BGP-A

PN W09729765-A1.
XX

PD 21-AUG-1997.
XX

PF 13-FEB-1997; 97WO-US02218.
XX

PR 16-FEB-1996; 96US-0011834.
XX

PA (REGC) UNIV CALIFORNIA.
XX

PI selsted ME;
XX

DR WPT; 1997-424753/39.
DR P-PSDB; AAW23722.

Antimicrobial (poly)peptide(s) f

useful therapeutically, as present treatment and in agriculture

XX
PS
v v

Claim 13; Fig 4; 56pp; English.

This cDNA sequence encodes the p

CC claimed antimicrobial peptide from
CC bovine granulocyte peptide A or
CC 100-143, 5

isolated from bovine bone marrow procedures. The encoded BGP-A p

cc signal peptide, a 156-residue pr

CC acids corresponding to mature BGP-A. BGP-A and the murine
CC homologue, MGP-A (see AAM27325), exhibit activity against
CC Gram-positive and Gram-negative bacteria, fungi and viruses,
CC specifically *Staphylococcus aureus*, *Escherichia coli*, *Candida*
CC *albicans*, *Salmonella typhimurium* and *C. neoformans* (claimed).
CC They can be used in human or veterinary medicine (particularly to
CC treat disorders associated with lipopolysaccharides, e.g. sepsis
CC and endotoxaemia) or as preservatives in food products or in water
CC supplies (claimed). They can also be applied to crops to reduce
CC post-harvest spoilage or expressed in transgenic plants to increase
CC their disease resistance. They have low immunogenicity.
XX
XX Sequence 608 BP; 126 A; 234 C; 201 G; 127 T; 0 other

Query Match 51.8%; Score 284.2; DB 18;
Best Local Similarity 71.6%; Pred. No. 6.4e-76;
Matches 389; Conservative 0; Mismatches 148; Indels 6

QY	6	gtttgcctgtgctctctctgcctcctcctgggtct-----ggcaacctcctcgcagtttcat	59
DB	49	gctgcctgggtctctcctcctcctcctggcctcggggcggtccaagatcgcgcgacat	108
QY	60	cgtgcgccgcagtgagtggaagccctgcactccgagtgctctagccgctcgtggggcacc	119
DB	109	cgtgtcccgcggaagtgggggccctggcatacaagtgcagcaggtgaagacagcc	168
QY	120	agttcgctacgttggtgatctcacacacagcggcgagcttctcacaacgcccgagctcctg	179
DB	169	tgtgcgtacgttggtgtctgcacacgcgcgagcgtctgcacaactcggcctcgtg	228
QY	180	tgaacagcagcccgcaatgtgcagcattaccacaagaatgagctgggtggtgcgatgt	239
DB	229	ccagaggcagggcccaaaacgtgcagtactaccacgtgcgggagcggggtctggtcgacgt	288
QY	240	agctacaactccttatgtgagagcaggtctatgctctatgaagccgagcgtcggaacat	299
DB	289	gggtacaatttctcctgctcgtggaagatgggtcgtgtatgaggccgggctgggaacac	348
QY	300	caaggggtgaccacacagggcccctctcggaatcccatgctctattggcatcaccttcattgg	359
DB	349	cttaggtgtcaactctgggcccacgtggaaccccatagccatcggcatctccttcattgg	408
QY	360	gaacttcattgaccgggtaccgcccaagcggggccctcctgctgcctcaattctctgga	419
DB	409	caactacatgcattcggtgccccggcctctgctctcagggggcccgagagctgctggc	468
QY	420	atgtggggtgtctcggggcttcctgagatcccaactatgaagtcacaaggacacccggatgt	479
DB	469	tgtggtgcagctcggggatcacctgactccttaactcgaagtcaaggacacccggatgt	528
QY	480	gcacaagcactctctccaggtgaccacaactctatcaggttcattcccaagctgggaacacta	539
DB	529	gcagcagacgctctctccaggggacagagctctataaaatcatccagcagtggccgcacta	588
QY	540	ccg 542	
DB	589	ccg 591	

RESULT 10

AAZ42549
ID AAZ42549

XX
AC
AAZ425XX
DT 01-FEB-1964

XX	Human
----	-------

XX	Human
----	-------

KW gene
KW foren:

regulation; identification; ss.

XX Homo sapiens.

XX WO9953051-A2.

XX 21-OCT-1999.

XX 09-APR-1999; 99WO-IB00712.

XX 09-APR-1998; 98US-0057719.

XX 28-APR-1998; 98US-0069047.

XX (GEST) GENSET.

XX Dumas Milne Edwards J, Duclert A, Giordano J;

XX WPI; 2000-038446/03.

XX P-PSDB; AAY64935.

Novel secreted protein 5' expressed sequence tag sequences used in diagnostic, forensic, gene therapy, and chromosome mapping procedures

Claim 1; Page 336; 837pp; English.

XX AA42265 to AA43075 represent novel 5' expressed sequence tag (EST) sequences, corresponding to human secreted proteins. AAY64651 to AAY65438 represent the EST-related proteins corresponding to AA42265 to AA43052. The 5' ESTs can be used for producing secreted human gene products. They can be used to identify and isolate 5' untranslated regions (UTRs) and upstream regulatory regions which control the location, development stage, rate, and quantity of protein synthesis, as well as stability of mRNA. The ESTs are also useful as probes for chromosome mapping, and to obtain full length cDNA clones. The ESTs can also be used in forensic procedures to identify individuals, or in diagnostic procedures to identify individuals having genetic diseases resulting from abnormal gene expression. The products may also be used in gene therapy protocols. The nucleic acids encoding signal peptides can be used for directing extracellular secretion of a polypeptide or the insertion of a polypeptide into a membrane, or importing a polypeptide into a cell. The proteins encoded by the EST sequences may be useful in treating a variety of human conditions. Secreted proteins have therapeutic value, and the identification of new secreted proteins is valuable. AA42249 to AA42264 and AAY64644 to AAY64650 represent sequences used in the exemplification of the present invention.

XX Sequence 380 BP; 67 A; 130 C; 114 G; 63 T; 6 other;

Query Match 28.3%; Score 155.6; DB 21; Length 380;

Best Local Similarity 75.1%; Pred. No. 2.9e-37;

Matches 199; Conservative 5; Mismatches 60; Indels 1; Gaps 1;

QY 44 cctcctgagttctatcgtgcccgagtcagtgagggccctgacacagcggcagcttctgca 103
||| ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 115 cctgctcagcccatagtcgcccggagagtggtggaagccctggaacagagtcgccc 174
||| ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 104 gcgcctggggcaccagttcgtacgttggtgatctcacacacagcggcagcttctgca 163
||| ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 175 agcacctgagctgcccctacgtatggtggtatcgacacagcggcgagcagctgca 234
||| ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 164 acagcccgagactctgaaacagacagcggcccgcaatgtgcagcattacacagaatgagc 223
||| ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 235 acacccgcctctgcccagcagcggcccggaatgtgcagcactacacatgaagacac 294
||| ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 224 tgggctggtgcagctagcctcacactccctt-attgagagcggctcagctctatgaa 282
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 295 tgggctggtgcagctggtgctacactcctngattggagagcggctcgtatcagrag 354
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 283 ggcgcagctggaacatacaggtg 307
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 355 ggcgcgtggtggaacttaccgggts 379

RESULT 11

AA04007

ID AAD04007 standard; cDNA; 1107 BP.

XX

AC AAD04007;

XX

DT 02-JUL-2001 (first entry)

XX

DE Human full length Zgpai cDNA #2 degenerate sequence.

XX

KW Human; granulocyte peptide A; GP-A; Zgpai; cytostatic; antiinflammatory; vulnery; dermatological; anti-microbial; gastrointestinal disease; pulmonary; dental carries; periodontal disease; gene therapy; AIDS; Acquired Immune Deficiency Syndrome; epithelial wound; cystic fibrosis; infection; urinary tract; respiratory; vaginal; lung; skin; cancer; ovarian; rectal; chromosome 1; ss.

XX

OS Homo sapiens.

XX

PN WO200129224-A2.

XX

PD 26-APR-2001.

XX

PF 20-OCT-2000; 2000WO-US29177.

XX

PR 20-OCT-1999; 99US-0160712.

PR 12-JUL-2000; 2000US-0218070.

XX

PA (ZYMO) ZYMOGENETICS INC.

XX

PI Conklin DC, Adler DA, Fox BA;

XX

DR WPI; 2001-290918/30.

XX

PT New granulocyte peptide homolog, zgpai polypeptide, for research applications, diagnosis and treatment of cancer, periodontal, gastrointestinal disease, urinary tract, skin and lung infections

PT

PS Claim 3; Page 111; 114pp; English.

XX

CC This sequence is stated as encoding human truncated granulocyte peptide (GP-A) homolog, zgpai protein. This sequence is a degenerate version of the coding region of SEQ ID NO: 7 shown in AAD04006. Zgpai gene is located on human chromosome 1. Zgpai polypeptides are useful for producing antibodies which are useful for detecting cancer. Zgpai polypeptides having anti-microbial activity are useful for treating dental carries, periodontal disease, thrush, gastrointestinal diseases, urinary tract infections, respiratory infections, vaginal infections, acquired immune deficiency syndrome (AIDS) and lung infections associated with cystic fibrosis and prevention of infection in skin and other epithelial wounds. Zgpai-cytokine fusion proteins are useful for enhancing in vivo killing of target tissues (epithelial cancers, and more specifically lung, ovarian and rectal cancers). Zgpai polypeptides, fragments, fusion proteins or agonists are useful in in vitro studies of exogenous microorganism infections such as bacterial, viral or fungal infection and also to study epithelial cell defensin induction in cell culture. Zgpai antibodies, polynucleotides and polypeptides are useful for detection of zgpai polypeptide, mRNA or anti-zgpai antibodies, thus serving as markers for detecting genetic diseases or cancers. Zgpai sequences are useful as diagnostics in forensic DNA profiling and as probes or primers to clone 5' non-coding regions of a zgpai gene. Zgpai antibodies are useful for tagging cells that express zgpai, for screening expression libraries and as neutralizing antibodies or as antagonists to block zgpai activity in vitro and in vivo. Zgpai gene is also useful in gene therapy.

XX

SQ Sequence 1107 BP; 186 A; 117 C; 201 G; 144 T; 459 other;

Query Match 23.1%; Score 126.8; DB 22; Length 1107;

Best Local Similarity 38.1%; Pred. No. 2.2e-28;

Matches 192; Conservative 76; Mismatches 230; Indels 6; Gaps 2;

Search completed: December 17, 2001, 09:35:26
Job time: 6376 sec

GenCore version 4.5
 Copyright (c) 1993 - 2000 CompuGen Ltd.
 OM protein - protein search, using sw model
 Run on: December 17, 2001, 07:49:38 ; Search time 22.96 Seconds
 (without alignments)
 133.786 Million cell updates/sec

Title: US-09-462-625-2_COPY_55_75
 Perfect score: 118
 Sequence: 1 NSPDSCEQQARNVQHYHKNEL 21

Scoring table: BLOSUM62
 Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues
 Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0
 Maximum DB seq length: 2000000000
 Post-processing: Minimum Match 0%
 Maximum Match 100%
 Listing first 150 summaries

Database : SPTREMBL17:*
 1: sp_archaea:*
 2: sp_bacteria:*
 3: sp_fungi:*
 4: sp_human:*
 5: sp_invertebrate:*
 6: sp_mammal:*
 7: sp_mhc:*
 8: sp_organelle:*
 9: sp_phage:*
 10: sp_plant:*
 11: sp_rodent:*
 12: sp_virus:*
 13: sp_vertebrate:*
 14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES			
Result No.	Score	Query Match Length	Description
1	118	100.0	182 11 088593
2	86	72.9	196 4 075594
3	79	66.9	183 11 09JLN4
4	65	55.1	193 6 09GK12
5	52	44.1	182 5 076537
6	50	42.4	440 6 09GLC6
7	48.5	41.1	523 5 018378
8	48	40.7	203 5 09VYX7
9	47	39.8	696 5 09V7M0
10	46	39.0	402 5 09VDQ4
11	46	39.0	502 4 09HCT9
12	46	39.0	502 4 09HCT9
13	45	38.1	139 10 093795
14	45	38.1	143 2 031779
15	45	38.1	329 5 09VEY8
16	45	38.1	604 2 09RYI1
17	44	37.3	185 4 09V3R1
18	44	37.3	185 4 09U114
19	44	37.3	185 11 035394

20	44	37.3	185	11	Q920S9	Q920S9 mus musculus
21	44	37.3	227	10	Q405I3	Q405I3 nicotiana t
22	44	37.3	288	5	Q917E1	Q917E1 drosophila
23	44	37.3	358	5	Q23222	Q23222 caenorhabdi
24	44	37.3	412	5	Q9V325	Q9V325 drosophila
25	44	37.3	417	2	Q9ZFG5	Q9ZFG5 staphylococ
26	44	37.3	439	10	Q9FJP7	Q9FJP7 arabidopsis
27	44	37.3	640	5	Q24443	Q24443 drosophila
28	44	37.3	701	5	Q9GN15	Q9GN15 leishmania
29	44	37.3	734	5	Q9XTT2	Q9XTT2 caenorhabdi
30	44	37.3	772	10	Q81491	Q81491 arabidopsis
31	44	37.3	778	10	Q9MAV5	Q9MAV5 arabidopsis
32	44	37.3	792	10	P92943	P92943 arabidopsis
33	44	37.3	945	2	Q59999	Q59999 synecocyst
34	44	37.3	1057	10	Q9FDZ3	Q9FDZ3 arabidopsis
35	44	37.3	1057	10	Q9C5T9	Q9C5T9 arabidopsis
36	44	37.3	1080	10	Q9C5U0	Q9C5U0 arabidopsis
37	44	37.3	1080	10	Q9C5T8	Q9C5T8 arabidopsis
38	44	37.3	139	10	Q93787	Q93787 gossypium h
39	43	36.4	170	10	Q81055	Q81055 arabidopsis
40	43	36.4	266	11	Q9D9J2	Q9D9J2 mus musculu
41	43	36.4	319	11	O55014	O55014 mus musculu
42	43	36.4	319	11	Q9Z1F0	Q9Z1F0 rattus norv
43	43	36.4	319	11	Q9CR42	Q9CR42 mus musculu
44	43	36.4	326	2	Q9XC26	Q9XC26 acinetobact
45	43	36.4	327	5	Q94244	Q94244 caenorhabdi
46	43	36.4	347	2	O51587	O51587 borrelia bu
47	43	36.4	353	5	Q9XZX7	Q9XZX7 leishmania
48	43	36.4	448	12	O65153	O65153 african swi
49	43	36.4	452	2	Q9Z5G6	Q9Z5G6 mycobacteri
50	43	36.4	743	5	Q9V5J4	Q9V5J4 drosophila
51	43	36.4	922	10	Q9FSE5	Q9FSE5 oryza sativ
52	42.5	36.0	743	12	Q83111	Q83111 mouse adeno
53	42	35.6	65	11	Q9Z0Q1	Q9Z0Q1 mus musculu
54	42	35.6	65	11	Q9WUT0	Q9WUT0 mus musculu
55	42	35.6	74	11	Q9Z0Q0	Q9Z0Q0 mus musculu
56	42	35.6	74	11	Q9WUT1	Q9WUT1 mus musculu
57	42	35.6	96	11	Q9M21	Q9M21 mus musculu
58	42	35.6	184	5	Q9V4X2	Q9V4X2 drosophila
59	42	35.6	202	10	Q9M8Y9	Q9M8Y9 arabidopsis
60	42	35.6	222	10	Q40171	Q40171 lycopersico
61	42	35.6	288	2	Q9RYT6	Q9RYT6 deinococcus
62	42	35.6	324	10	Q9M4G1	Q9M4G1 solanum tub
63	42	35.6	408	5	O01634	O01634 caenorhabdi
64	42	35.6	442	2	Q9PJ88	Q9PJ88 campylobact
65	42	35.6	470	5	Q9W249	Q9W249 drosophila
66	42	35.6	498	5	P91527	P91527 caenorhabdi
67	42	35.6	501	11	Q9WV38	Q9WV38 mus musculu
68	42	35.6	507	12	Q9DK06	Q9DK06 allpahuayo
69	42	35.6	507	12	Q9DK03	Q9DK03 allpahuayo
70	42	35.6	508	12	O90423	O90423 pichinde ar
71	42	35.6	508	12	Q9YTX1	Q9YTX1 pichinde ar
72	42	35.6	508	12	Q9YTW9	Q9YTW9 pichinde ar
73	42	35.6	508	12	Q9YTW8	Q9YTW8 pichinde ar
74	42	35.6	508	12	O11997	O11997 pichinde ar
75	42	35.6	508	12	O11998	O11998 pichinde ar
76	42	35.6	509	12	O11999	O11999 pichinde ar
77	42	35.6	545	10	Q9SFP31	Q9SFP31 arabidopsis
78	42	35.6	789	5	Q9GZ73	Q9GZ73 drosophila
79	42	35.6	816	10	Q9SJT3	Q9SJT3 arabidopsis
80	42	35.6	855	5	Q9VTU0	Q9VTU0 drosophila
81	42	35.6	884	5	O61543	O61543 drosophila
82	42	35.6	884	5	Q9VEV9	Q9VEV9 drosophila
83	42	35.6	1104	4	O60460	O60460 homo sapien
84	42	35.6	1373	4	O75372	O75372 homo sapien
85	42	35.6	3191	5	O01335	O01335 caenorhabdi
86	41.5	35.2	337	2	O25602	O25602 helicobacte
87	41.5	35.2	337	2	Q9ZKQ3	Q9ZKQ3 helicobacte
88	41.5	35.2	463	10	Q9SKE4	Q9SKE4 arabidopsis
89	41.5	35.2	692	5	O45101	O45101 caenorhabdi
90	41.5	35.2	1029	10	Q9SYX2	Q9SYX2 arabidopsis
91	41	34.7	88	2	Q9K2P2	Q9K2P2 vibrio chol
92	41	34.7	146	2	Q9KAB8	Q9KAB8 bacillus ha

RESULT	1	ALIGNMENTS	PRELIMINARY;	PRT;	182 AA.
ID	088593				
AC	088593; 062185;				
DT	01-NOV-1998 (TReMBLrel. 08, Created)				
DT	01-NOV-1998 (TReMBLrel. 08, Last sequence update)				
DT	01-JUN-2001 (TReMBLrel. 17, Last annotation update)				
DE	PEPTIDOGLYCAN RECOGNITION PROTEIN PRECURSOR.				
GN	PGLYRP OR PGRP OR TAG7.				
OS	Mus musculus (Mouse).				
93	41	34.7	216	5	Q9X254
94	41	34.7	216	5	Q9V3V2
95	41	34.7	382	2	Q9ETP3
96	41	34.7	430	10	Q9AW73
97	41	34.7	459	3	P78720
98	41	34.7	471	10	Q9AWR2
99	41	34.7	474	10	Q48591
100	41	34.7	506	8	Q47143
101	41	34.7	506	8	Q47146
102	41	34.7	507	8	Q47122
103	41	34.7	512	13	Q9PSM0
104	41	34.7	515	10	Q48592
105	41	34.7	575	10	Q39117
106	41	34.7	589	3	Q9USP9
107	41	34.7	594	10	Q48590
108	41	34.7	594	10	Q9C882
109	41	34.7	730	10	Q9LP25
110	41	34.7	929	5	Q17548
111	41	34.7	1338	5	Q9V416
112	41	34.7	1778	5	Q9VLD8
113	41	34.7	1948	1	Q28189
114	41	34.7	2035	5	Q9GS19
115	40.5	34.3	327	2	Q9F0B8
116	40.5	34.3	430	5	Q44725
117	40.5	34.3	849	5	Q15984
118	40.5	34.3	1341	5	Q9GRK2
119	40	33.9	133	3	Q13577
120	40	33.9	147	2	Q9KAE9
121	40	33.9	147	10	Q22831
122	40	33.9	178	12	Q71251
123	40	33.9	195	5	Q97369
124	40	33.9	202	10	Q65136
125	40	33.9	225	5	Q27014
126	40	33.9	225	5	Q27013
127	40	33.9	226	10	Q9FGU7
128	40	33.9	319	4	Q15327
129	40	33.9	319	6	Q9TU71
130	40	33.9	340	1	Q9HJDI
131	40	33.9	359	10	Q9LUF4
132	40	33.9	379	2	Q9X2C7
133	40	33.9	389	2	Q9PGI7
134	40	33.9	393	5	Q9XU07
135	40	33.9	393	5	Q9PR55
136	40	33.9	398	2	Q914M8
137	40	33.9	419	11	Q9D4K7
138	40	33.9	422	2	P95734
139	40	33.9	500	11	Q9CWW5
140	40	33.9	508	5	Q917F4
141	40	33.9	509	12	Q995C5
142	40	33.9	531	2	Q9KU64
143	40	33.9	539	3	Q14000
144	40	33.9	547	5	Q9N3E8
145	40	33.9	574	3	Q93990
146	40	33.9	595	4	Q75178
147	40	33.9	646	10	Q9FT80
148	40	33.9	649	2	Q9ZFY0
149	40	33.9	692	4	Q95064
150	40	33.9	695	10	Q9FFX6

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus; NCBI_TaxID=10090;

[1] SEQUENCE FROM N.A., FUNCTION, AND TISSUE SPECIFICITY. TISSUE=SPLEEN; MEDLINE=98374308; PubMed=9707603; Kang D., Liu G., Lundstroem A., Gellius E., Steiner H.; "A peptidoglycan recognition protein in innate immunity conserved from insects to humans."; Proc. Natl. Acad. Sci. U.S.A. 95:10078-10082(1998).

[2] SEQUENCE FROM N.A., FUNCTION, AND SUBCELLULAR LOCATION. MEDLINE=98325081; PubMed=9660837; Kiselev S.L., Kustikova O.S., Korobko E.V., Prokhortchouk E.B., Kabishev A.A., Lukanidin E.M., Georgiev G.P.; "Molecular cloning and characterization of the mouse tag7 gene encoding a novel cytokine."; J. Biol. Chem. 273:18633-18639(1998).

[3] SEQUENCE FROM N.A. Stayton W.B., Rigaa A., Hancock J.D., Zaugg J.K., Le T.V., Trautman M.S., Spangrude G.J., Carroll W.L., Schibler K.R.; "Granulocyte-colony stimulating factor up-regulates expression of murine tag7 during myeloid differentiation."; Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.

[4] SEQUENCE FROM N.A. STRAIN=C57BL/6J;

Query Match 72.9%; Score 86; DB 4; Length 196;
Best Local Similarity 71.4%; Pred. No. 3.1e-06;
Matches 15; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

OY 1 NSPDSCEQQAARNVQHYNKEL 21
I:-|::|:||||||| |
DB 68 NTFASCCQQAARNVQHYNMKTLL 88

RESULT 3
OYJLNA PRELIMINARY; PRT; 183 AA.
AC OYJLN4:
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DE 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE PEPTIDOGLYCAN RECOGNITION PROTEIN PCRP.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_Taxid=101116;
RN [1]
FP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=SPLEEN;
RA Rehman A., Teodecki E.E., Krueger J.M.;
RL Submitted (MAY-1999) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AF154114; AAF73252.1;
SQ SEQUENCE 183 AA; 20590 MW; 5B9C1B7AA8A2EC21 CRC64;

Query Match 66.9%; Score 79; DB 11; Length 183;
Best Local Similarity 71.4%; Pred. No. 3.8e-05;
Matches 15; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

OY 1 NSPDSCEQQAARNVQHYNKEL 21
:||||||| | :|
DB 56 SSPDSCQQAARNVOLYQMQL 76

RESULT 4
OYGKL2 PRELIMINARY; PRT; 193 AA.
ID OYGKL2:
AC OYGKL2:
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-NAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-NAR-2001 (TrEMBLrel. 16, Last annotation update)
DE PEPTIDOGLYCAN RECOGNITION PROTEIN PRECURSOR.
OS Camelus dromedarius (Dromedary) (Arabian camel).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Tylopoda; Camelidae; Camelus.
OX NCBI_Taxid=9838;
RN [1]
FP SEQUENCE FROM N.A.
RC TISSUE=LACTATING MAMMARY GLAND;
RA Kappeler S.R., Farah Z., Fuhan Z.;
RL "Milk as a Source of Camel (Camelus dromedarius) Peptidoglycan Recognition Protein.";
DE Submitted (DEC-1998) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AJ131676; CAC19553.1;
KW Signal.
FT SIGNAL.
FT CHAIN
SQ SEQUENCE 193 AA; 21377 MW; B6A1BD818030A7CB CRC64;

Query Match 55.1%; Score 65; DB 6; Length 193;
Best Local Similarity 57.1%; Pred. No. 0.0073;
Matches 12; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

OY 1 NSPDSCEQQAARNVQHYNKEL 21

RESULT 9
Q9V7M0 Q9V7M0 PRELIMINARY; PRT; 696 AA.
Q9V7M0;
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DT 01-MAY-2000 (TReMBLrel. 13, Last annotation update)
DE CG15702 PROTEIN.
CG15702.
GN Drosophila melanogaster (Fruit fly).
OS Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OC NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY.
RX MEDLINE=20196006; PubMed=107311132;
RA Adams M.D., Celnikier S.E., Holt R.A., Evans C.A., Cocayne J.D.,
RA Amaratides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Vandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazeg R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brockstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieux E., Center A., Chandra I.,
RA Cherry J.M., Cavelli S., Dahke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrieria S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodgate T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RA "The genome sequence of Drosophila melanogaster.";
RRL scfence 287:1185-2195(2000).
RRT EMBL: AE003807; AAF58028.1;
RDR FlyBase; FBgn0034101; CG15702;
SQ SEQUENCE 696 AA; 79755 MW; 7BC2620B01FAE816 CRC64;

```

Query Match          39.8%; Score 47; DB 5; Length 696;
Best Local Similarity 40.0%; Pred. No. 19;
Matches 8; Conservative 4; Mismatches 8; Indels 0; Gaps 0;

2y 2 SPDSCSEQAARNVQHYHKNEL 21
   :|| | | :||| ||
db 61 TPDQVELQVLGLRHYYSKEL 80

```

RA	Cherry J.M., Cawley S., Dahlike C., Davenport L.B., Davies P., de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M., Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P., Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W., Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K., Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M., Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J., Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C., Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A., Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z., Lasko P., Lei Y., Levitsky A.C., Li J., Li Z., Liang Y., Lin X., Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D., Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A., Mount S.M., Mov M., Murphy B., Murphy L., Muzny D.M., Nelson D.L., Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M., Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G., Reinart K., Remington K., Saunders R.D.C., Scheeler F., Shen H., Shue B.C., Sidon-Kiamos I., Simpson M., Skupski M.P., Smith T., Spier E.C., Spradling A.C., Stapleton M., Strong R., Sun E., Svirskaas R., Tector C., Turner R., Venter E., Wang A.H., Wang X., Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J., Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A., Ye J., Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L., Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O., Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.; RT "The genome sequence of Drosophila melanogaster"; RL Science 287:2185-2195(2000). DR EMBL: AE003728; AAF55736.1; -. DR HSSP; P25490; IUBD. DR DR FlyBase; FBgn0038767; CG4413. DR InterPro; IPRO00822; ZnF-C2H2. DR Pfam; PF00096; zf-C2H2; 3. DR SMART; SM00355; ZnF_C2H2; 3. DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 3. DR PROSITE; PS0157; ZINC_FINGER_C2H2_2; 3. KW DNA-binding; Metal-binding; Zinc-finger. SQ SEQUENCE 402 AA; 45857 MW; 6B1D1D6AA353830A CRC64;					
QY	Query Match Best Local Similarity Matches	39.0%; 41.2%; Conservative	Score 46; DB 5; Pred. No. 17;	Length 402; Mismatches	7; Indels	0; Gaps
Db	3 PDCEQQQAARNVQHYNK 19 I::: 234 PEKCDRSGRKPVAHYKN 250					
RESULT 12						
Q9HCT9	ID	PRELIMINARY;	PRT;	502 AA.		
AC	O9HCT9;					
DT	01-MAR-2001 (TREMBLrel. 16, Created)					
DT	01-WAR-2001 (TREMBLrel. 16, Last sequence update)					
DT	01-JUN-2001 (TREMBLrel. 17, Last annotation update)					
GN	TUMOR ENDOTHELIAL MARKER 7 PRECURSOR.					
OS	Homo sapiens (Human).					
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
OX	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.					
NCBI_TaxID=9606;	[1]					
SEQUENCE FROM N.A.						
MEDLINE=20407466; PubMed=10947988;						
St Croix B., Rago C., Velculescu V., Traverso G., Romans K.E., Montgomery E., Lai A., Riggs G.J., Lengauer C., Vogelstein B., Kinzler K.W.;						
"Genes expressed in human tumor endothelium.";						
Science 289:1197-1202(2000).						
EMBL; AF279144; AAG00869.1; -.						
InterPro; IPR003886; Nidogen_ext.						
InterPro; IPR002165; Plexin_repeat.						
InterPro; IPR003659; PSI.						

DR Pfam; PF01437; plexin-repeat; 1.
 DR SMART; SM00539; NIDO; 1.
 DR SMART; SM00423; PSI; 1.
 KW SIGNAL.
 FT CHAIN 1 18 POTENTIAL.
 FT CHAIN 19 502 TUMOR ENDOTHELIAL MARKER 7.
 SQ SEQUENCE 502 AA; 55647 MW; 94E0A8E15B296144 CRC64;

Query Match 39.0%; Score 46; DB 4; Length 502;
 Best Local Similarity 45.0%; Pred. No. 21;
 Matches 9; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

OY 2 SPDSCEQOQARNVQVHKNE 21
 ||| | : : : | : | : |

Db 266 SPDVPESRRSIFVHRLEL 285

RESULT 13

Q39795 ID Q39795 PRELIMINARY; PRT; 139 AA.

DT 01-NOV-1996 (TREMBLrel. 01, Created)

DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)

DE 01-JUN-2001 (TREMBLrel. 17, Last annotation update)

DE 2S ALBUMIN STORAGE PROTEIN PRECURSOR.

GN MAY5-A.

OS Gossypium hirsutum (Upland cotton).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;

OC eucosids II; Malvales; Malvaceae; Gossypium.

OX NCBI_TaxID=3635;

RT [1]

RL SEQUENCE FROM N.A.

RA Galau G.A.; Wang H.Y.C.; Hughes D.W.;

RL Submitted (JAN-1992) to the EMBL/GenBank/DBJ databases.

DR EMBL; M86213; AAA33066.1; -

DR Mendel; 8406; Goshi; 2374; 8406.

DR InterPro; IPR001769; Cereal_tryp_amyl_inh.

DR InterPro; IPR003612; AAI.

DR Pfam; PF00234; tryp_alpha_amyl; 1.

DR SMART; SM00499; AAI; 1.

KW SIGNAL.

FT SIGNAL 1 20 POTENTIAL.

FT CHAIN 33 59 2S ALBUMIN STORAGE PROTEIN.

FT CHAIN 64 139 2S ALBUMIN STORAGE PROTEIN.

SQ SEQUENCE 139 AA; 15700 MW; 02ACE24FFEC9EF90 CRC64;

Query Match 38.1%; Score 45; DB 10; Length 139;
 Best Local Similarity 45.8%; Pred. No. 9;
 Matches 11; Conservative 1; Mismatches 6; Indels 6; Gaps 1;

OY 4 DSCEQOQARNVQVHKNE 21
 ||||| | : : : | : | : |

Db 34 DSCEQOQARNVQVHKNE 57

RESULT 14

O31779 ID O31779 PRELIMINARY; PRT; 143 AA.

AC O31779

DT 01-JAN-1998 (TREMBLrel. 05, Created)

DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)

DE 01-NOV-1998 (TREMBLrel. 08, Last annotation update)

DE YMCA PROTEIN.

GN YMCA.

OS Bacillus subtilis.

OC Bacteria; Firmicutes; Bacillus/Clostridium group;

OC Bacillus/Staphylococcus group; Bacillus.

OX NCBI_TaxID=1423;

RT [1]

RL SEQUENCE FROM N.A.

RC STRAIN=168;
 RX MEDLINE=98044033; PubMed=9384377;
 RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
 RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
 RA Borris R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
 RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
 RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
 RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
 RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
 RA Fritz C., Fujita M., Fujita Y., Fuma S., Gallazzi A., Galleron N.,
 RA Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
 RA Hilbert H., Holsappel S., Hosono S., Hulio M.F., Itaya M., Jones L.,
 RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
 RA Kobayashi Y., Koetter P., Koningstein G., Krogh S., Kumano M.,
 RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
 RA Lee S.M., Levine A., Liu H., Masuda S., Maue J.C., Medigue C.,
 RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
 RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
 RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
 RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
 RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
 RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
 RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
 RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
 RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
 RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
 RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenecker T.,
 RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
 RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.,
 RT "The complete genome sequence of the gram-positive bacterium Bacillus
 RT subtilis.";
 RL Nature 390:249-256(1997).
 RL [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RA Kunst F., Ogasawara N., Yoshikawa H., Danchin A.;
 RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL; 299112; CAB13575.1; -
 KW Complete proteome.
 SQ SEQUENCE 143 AA; 16166 MW; 750A24D99B463C25 CRC64;

Query Match 38.1%; Score 45; DB 2; Length 143;
 Best Local Similarity 40.0%; Pred. No. 9.2;
 Matches 8; Conservative 6; Mismatches 6; Indels 0; Gaps 0;

OY 1 NSPDSCEQOQARNVQVHKNE 20
 | : : : | : : : | : |

Db 47 NQIKALQKQAVNLKHYEKE 66

RESULT 15

O9VFY8 ID O9VFY8 PRELIMINARY; PRT; 329 AA.

AC O9VFY8

DT 01-MAY-2000 (TREMBLrel. 13, Created)

DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)

DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)

DE CG10148 PROTEIN.

GN CG10148

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI_TaxID=7227;

RT [1]

RL SEQUENCE FROM N.A.

RC STRAIN=BERKELEY;

RX MEDLINE=20196006; PubMed=10731132;

RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,

RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galie R.F.,

RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,


```

OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=HIPOTHALAMUS;
RA Jin W., Huang C., Wu T., Peng Y., Gu Y., Zhang L., Jiang C., Li Y.,
RA Han Z., Wang Y., Chen Z., Fu G.,
RT "A novel gene expressed in human hypothalamus.";
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF112202; AAF1790.1; -
SQ SEQUENCE 185 AA; 20648 MW; A85BD5BDCF6C23E4 CRC64;

Query Match
Best Local Similarity 37.3%; Score 44; DB 4; Length 185;
Matches 7; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

QY 1 NSPDSCEQOARNVQVYHKN 19
DB 59 NVGELCQRLVRNVYQSN 77
| : : : | | | : |
| : : : | | | : |

RESULT 19
O35394 PRELIMINARY; PRT; 185 AA.
AC O35394;
DT 01-JAN-1998 (TReMBLrel. 05, Created)
DT 01-JAN-1998 (TReMBLrel. 05, Last sequence update)
DE PRENYLATED RAB ACCEPTOR 1.
GN PRAL.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RA Martincic I., Peralta M.E., Ngsee J.K.;
RL J. Biol. Chem. 272:0-0(1997).
DR EMBL: AF025506; AAB81721.1; -
SQ SEQUENCE 185 AA; 20643 MW; DA6341AE66F5C2F0 CRC64;

Query Match
Best Local Similarity 37.3%; Score 44; DB 11; Length 185;
Matches 7; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

QY 1 NSPDSCEQOARNVQVYHKN 19
DB 59 NVGELCQRLVRNVYQSN 77
| : : : | | | : |
| : : : | | | : |

RESULT 20
Q920S9 PRELIMINARY; PRT; 185 AA.
ID Q920S9;
AC Q920S9;
DT 01-MAY-1999 (TReMBLrel. 10, Created)
DT 01-MAY-1999 (TReMBLrel. 10, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE PRENYLATED RAB ACCEPTOR 1 (2310040106RIK PROTEIN).
GN RABAC1 OR PRAL OR PRA OR 2310040106RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Behrend E.N., Kempainen R.J.;
RT "Specific interaction of Dexas with prenylated rab acceptor 1
(PRAL).";
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Li G.;

```

```

RT "Mouse prenylated Rab acceptor (mpRA).";
RN [3]
RP Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
RC SEQUENCE FROM N.A.
RX STRAIN=C57BL/6J; TISSUE=TONGUE, AND SMALL INTESTINE;
MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schirni L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai H., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-P.,
RA Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RN [1]
RP Nature 409:685-690(2001).
DR EMBL: AF120162; AAD17296.1; -
DR EMBL: AF252856; AAF68476.1; -
DR EMBL: AK009726; BAB26465.1; -
DR EMBL: AK008559; BAB25744.1; -
DR MGD: MGI:1201692; Rabac1.
DR MGD: MGI:1914946; 2310040106RIK.
SQ SEQUENCE 185 AA; 20619 MW; 86C82C0502B5CB5B CRC64;

Query Match
Best Local Similarity 37.3%; Score 44; DB 11; Length 185;
Matches 7; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

QY 1 NSPDSCEQOARNVQVYHKN 19
DB 59 NVGELCQRLVRNVYQSN 77
| : : : | | | : |
| : : : | | | : |

RESULT 21
Q40513 PRELIMINARY; PRT; 227 AA.
ID Q40513;
AC Q40513;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE MADS-BOX PROTEIN.
GN MADS.
OS Nicotiana tabacum (Common tobacco).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids I; Solanales; Solanaceae; Nicotiana.
OX NCBI_TaxID=4097;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97048712; PubMed=8893543;
RA Davies, D.I. Rosa, Eneva T., Saedler, Sommer;
RT "Alteration of tobacco floral organ identity by expression of
combinations of Antirrhinum MADS-box genes.";
RL Plant J. 10:663-677(1996).
CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -1- SIMILARITY: TO THE MADS DOMAIN FAMILY OF TRANSCRIPTION FACTORS.
DR EMBL: X96428; CAA65288.1; -
DR HSP: P11746; IMNN.
DR TRANSFAC; T03109; -
DR Mendel; 9260; Nicta; MADS; 9260.
DR InterPro; IPR002487; K-box.
DR InterPro; IPR002100; MADS-box.

```



```
DR Pfam; PF01486; K-box; 1.
DR Pfam; PF00319; SRF-TF; 1.
DR PRINTS; SM00404; MADSDOMAIN.
DR SMART; SM00432; MADS; 1.
DR PROSITE; PS00350; MADS_BOX_1; 1.
DR PROSITE; PS00066; MADS_BOX_2; 1.
KW DNA-binding; Nuclear protein; transcription regulation.
SQ SEQUENCE 227 AA; 26401 MW; FEF8A9A2AD7FEC4 CRC64;

Query Match 37.3%; Score 44; DB 10; Length 227;
Best Local Similarity 38.1%; Pred. No. 21;
Matches 8; Conservative 6; Mismatches 7; Indels 0; Gaps 0;

Qy 1 NSPDSCEQQAARNVOHYHKNEL 21
Db 145 NQIDTYKKVRNVEIHRNLL 165

RESULT 22
Q917E1 PRELIMINARY; PRT; 288 AA.
ID Q917E1
AC Q917E1
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE CG18853 PROTEIN.
GN CG18853
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Vandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champagne M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agyan A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew K.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrera S., Fleischmann W.,
RA Folsler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Laso P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazzolo M., Pitman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskaas R., Tector C., Turner E., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
```

```
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
DR EMBL; AE003839; AG22302.1; -.
DR FlyBase; FBgn0042173; CGI18853.
DR InterPro; IPR000474; DNA_photolyase.
DR InterPro; IPR000288; DNA_photolyase_2.
DR Pfam; PF00875; DNA_photolyase; 1.
DR ProDom; PD007711; DNA_photolyase_2; 1.
SQ SEQUENCE 288 AA; 33375 MW; 3682D8EBB635DDDD4 CRC64;
```

```
Query Match 37.3%; Score 44; DB 5; Length 288;
Best Local Similarity 33.3%; Pred. No. 26;
Matches 7; Conservative 6; Mismatches 8; Indels 0; Gaps 0;
```

```
Qy 1 NSPDSCEQQAARNVOHYHKNEL 21
Db 15 NEINNLRQQVRNLOHVQRKDI 35
```

```
RESULT 23
Q23222 PRELIMINARY; PRT; 358 AA.
ID Q23222
AC Q23222
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE W08D2.6 PROTEIN.
GN W08D2.6
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Peloderiinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Swinburne J., Ainscough R.;
RA Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
RX [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton A., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Showkeen R.,
RA Smalton N., Smith A., Sonnhammer E., Staden R., Sulston J.,
RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RL elegans";
RL Nature 368:32-38(1994).
DR EMBL; Z70271; CAA94234.1; -.
DR InterPro; IPR000087; Collagen.
DR InterPro; IPR002486; Col_cuticle_N.
DR Pfam; IPR003880; Phosphopant_attach.
DR Pfam; PF01391; Collagen; 2.
DR Pfam; PF01484; Col_cuticle_N; 1.
DR PROSITE; PS00012; PHOSPHOPANTHEINE; UNKNOWN1.
SQ SEQUENCE 358 AA; 36174 MW; E8276DC9946B0A56 CRC64;
```

```
Query Match 37.3%; Score 44; DB 5; Length 358;
Best Local Similarity 72.7%; Pred. No. 32;
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
```

```
Qy 2 SPDSCEQQAARN 12
Db 131 SPDSCEQQAARN 141
```

RESULT 24

Q9V325 Q9V325 PRELIMINARY; PRT; 412 AA.
 ID Q9V325; AC Q9V325;
 DT 01-MAY-2000 (TREMELrel. 13, Created)
 DT 01-MAY-2000 (TREMELrel. 13, Last sequence update)
 DT 01-JUN-2001 (TREMELrel. 17, Last annotation update)
 DE CG12822 PROTEIN.
 GN CG12822.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Ananadides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Randell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
 RA Brill J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
 RA Ballwe R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
 RA Curtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K.J., Evangelista C.C., Ferraz C., Fierstra S., Fleischmann W.,
 RA Foster C., Gong F., Gorrell J.H., Gu Z., Guan P., Glasser K.,
 RA Glodek A., Havel D., Harve D., Helman T.J., Hernandez J.R., Houch J.,
 RA Harris N.L., Harve D., Helman T.J., Hernandez J.R., Houch J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,
 RA Jalali M., Kalush F., Karpen G.H., Ye Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Laslo P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Moberly C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.L.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster."
 RL Science 287:2185-2195(2000).
 DR EMBL; AE003839; AAF59182.1;
 DR FLYBase; FBgn0033229; CG12822.
 DR InterPro; IPR001378; UPF0066.
 DR Pfam; PF01980; UPF0066; 1.
 DR ProDom; PD006705; UPF0066; 1.
 SQ SEQUENCE 412 AA; 46057 MW; AC030E0427132A8D CRC64;

Query Match 37.3%; Score 44; DB 5; Length 412;
 Best Local Similarity 33.3%; Pred. No. 36;
 Matches 7; Conservative 6; Mismatches 8; Indels 0; Gaps 0;

QY 1 NSPDSCEQQARNVQHYHKNEL 21
 | : : : | : : : | : : :
 Db 15 NEINNLQQRVRLQHVQKDI 35

RESULT 25

Q9ZFG5 Q9ZFG5 PRELIMINARY; PRT; 417 AA.
 ID Q9ZFG5; AC Q9ZFG5;
 DT 01-MAY-1999 (TREMELrel. 10, Created)
 DT 01-MAY-1999 (TREMELrel. 10, Last sequence update)
 DT 01-JUN-2001 (TREMELrel. 17, Last annotation update)
 DE FACTOR ESSENTIAL FOR METHICILLIN RESISTANCE FEMA.
 GN FEMA.
 OS Staphylococcus capitis.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Staphylococcus.
 OX NCBI_TaxID=29388;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC27840;
 RA Vannuffel P., Heusterspreute M., Bouyer M., Philippe M., Gala J.-L.;
 RT "Molecular characterization of femA from Staphylococcus hominis,
 RT Staphylococcus saprophyticus and Staphylococcus haemolyticus and femA-
 RT based discrimination of staphylococcal species.";
 RL Res. Microbiol. 0:0-0(1998).
 DR EMBL; AF099964; AAC69633.1;
 DR InterPro; IPR003447; FemAB.
 DR Pfam; PF02388; FemAB; 1.
 SQ SEQUENCE 417 AA; 49034 MW; C5B6F1BF30B62009 CRC64;

Query Match 37.3%; Score 44; DB 2; Length 417;
 Best Local Similarity 35.5%; Pred. No. 36;
 Matches 11; Conservative 4; Mismatches 6; Indels 10; Gaps 1;

QY 1 NSPDSCEQQ-----ARNVQHYHKNEL 21
 | : : : | : : : | : : :
 Db 281 NKRDNLQQLLDANQKIDEAKNLQOEHGNE 311

RESULT 26
 Q9FJF7 Q9FJF7 PRELIMINARY; PRT; 439 AA.
 ID Q9FJF7; AC Q9FJF7;
 DT 01-MAR-2001 (TREMELrel. 16, Created)
 DT 01-MAR-2001 (TREMELrel. 16, Last sequence update)
 DT 01-MAR-2001 (TREMELrel. 16, Last annotation update)
 DE GENOMIC DNA, CHROMOSOME 5, P1 CLONE:MQN23.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC Eurosid II; Brassicales; Brassicaceae; Arabidopsids.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=COLUMBIA;
 RX MEDLINE=98403884; PubMed=9734815;
 RA Kotani H., Nakamura Y., Sato S., Asamizu E., Kaneko T., Miyajima N.,
 RA Tabata S.;
 RT "Structural analysis of Arabidopsis thaliana chromosome 5. VI.
 RT Sequence features of the regions of 1,367,185 bp covered by 19
 RT physically assigned p1 and TAC clones.";
 RL DNA Res. 5:203-216(1998).
 DR EMBL; AB013395; BAB11654.1; -.
 SQ SEQUENCE 439 AA; 48708 MW; 0058A0EFD9CFC31E CRC64;

Query Match 37.3%; Score 44; DB 10; Length 439;
 Best Local Similarity 41.2%; Pred. No. 38;
 Matches 7; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

QY 1 NSPDSCEQQARNVQHYH 17
 | : : : | : : : | : : :
 Db 410 NQPNAAQQQAQQGQSFH 426

RESULT 27
Q24443


```

Submitted (NOV-1997) to the EMBL/GenBank/DBDJ databases.
[3]
RN RA SEQUENCE FROM N.A.
RA Ecker J.R.;
RN RA Submitted (DEC-1997) to the EMBL/GenBank/DBDJ databases.
[4]
RN RA SEQUENCE FROM N.A.
RA Ecker J.R.;
RN RA Submitted (JAN-1998) to the EMBL/GenBank/DBDJ databases.
[5]
RN RA SEQUENCE FROM N.A.
RA Ecker J.R.;
RN RA Submitted (APR-2000) to the EMBL/GenBank/DBDJ databases.
[6]
RN RA SEQUENCE FROM N.A.
RA Cheuk R., Shinn P., Brooks S., Buehler E., Chao Q., Johnson A.,
  Khan S., Kim C., Altafi H., Bei B., Chin C., Chiou J., Choi T.,
  Conn L., Conway A., Gonzalez A., Hansen N., Howing B., Koo T.,
  Lee J., Lenz C., Li J., Liu A., Liu J., Liu S., Mukharsy N.,
  Nguyen M., Palm C., Pham P., Sakano H., Schwartz J., Southwark
  Thaveri A., Toriumi M., Vaysberg M., Yu G., Davis R., Federspiel
  Theologis A., Ecker J.;
RN RA Submitted (MAY-2000) to the EMBL/GenBank/DBDJ databases.
RN RA EMBL; AC003113; AAF70851.1; -
  InterPro; IPR003864; DUF221.
  Pfam; PF02714; DUF221.1.
RN RA SEQUENCE 778 AA; 89308 MW; 52533674565B3E95 CRC64;

Query Match 37.3%; Score 44; DB 10; Length 778;
Best Local Similarity 34.3%; Pred. No. 65;
Matches 12; Conservative 2; Mismatches 5; Indels 1

QY 3 PDSECOQARNVQH-----YHKNEL 21
   ||| | | | | | | | | | | | | |
DB 210 PDSEDSISENVQHFFLVNHPDHYLTHQVVTNANEL 244

RESULT 32
PP2943
ID P92943 PRELIMINARY; PRT; 792 AA.
AC AC
DT 01-MAY-1997 (TrEMBLrel. 03, Created)
DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE CLC-D CHLORIDE CHANNEL PROTEIN.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheo-
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.
OC NCBI_TaxId=3702;;
RN [1]
RN RA SEQUENCE FROM N.A.
RN RC STRAIN-COLUMBIA; TISSUE-WHOLE PLANT;
RX MEDLINE-97126010; PubMed-8969232;
RA Hechenberger M., Schwappach B., Fischer W.N., Frommer W.B.,
RA Jentsch T.J., Steinmeyer K.;
RT "A family of putative chloride channels from Arabidopsis and
RT functional complementation of a yeast strain with a ClC gene
RT disruption.";
RN J. Biol. Chem. 271:33632-33638(1996).
RN EMBL; Z71450; CAA96065.1; -
RN Mendel; Z028; Arath:1417; Z028.
RN InterPro; IPR000644; CBS.
RN InterPro; IPR001807; Volt_CL_channel.
RN Pfam; PF00571; CBS; 2.
RN Pfam; PF00654; voltage_CLC; 1.
RN PRINTS; PR00762; CLCHANNEL.
RN SMART; SM00116; CBS; 2.
RN SEQUENCE 792 AA; 87047 MW; 45B0EDD2A8CCE85A CRC64;

Query Match 37.3%; Score 44; DB 10; Length 792;

```

```

Best Local Similarity 46.78; Pred. No. 66;
Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 1 NSPDCSQQARNVQH 15
    :|||:||||:|
Db 654 HSLPLCDPSARNIRH 668

RESULT 33
Q59999 PRELIMINARY; PRT; 945 AA.
AC Q59999;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DE CATION-TRANSPORTING ATPASE PACL.
GN PACL OR SLL0672.
OS Synechocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX NCBI_TaxID=1148;
RN
RP SEQUENCE FROM N.A.
RX MEDLINE=96127529; PubMed=8590279;
RA Kaneko T., Tanaka A., Sato S., Kotani H., Sazuka T., Miyajima N.,
RA Sugiura M., Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain PCC6803. I. Sequence features in the 1 Mb
RT region from map positions 64% to 92% of the genome.";
RL DNA Res. 2:153-166(1995).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=97061201; PubMed=8905231;
RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA Miyajima N., Hirotsawa M., Suglura M., Sasamoto S., Kimura T.,
RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K., Okumura S.,
RA Shimpo S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,
RA Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain PCC6803. II. Sequence determination of the
RT entire genome and assignment of potential protein-coding regions.";
RL DNA Res. 3:109-136(1996).
DR EMBL; D64005; BAA10744.1; -
DR HSSP; P04191; IEUL.
DR InterPro; IPR001757; E1-E2_ATPase.
DR InterPro; IPR001454; Hydrolase.
DR Pfam; PF001122; E1-E2_ATPase; 1.
DR Pfam; PF00702; Hydrolase; 1.
DR PRINTS; PR00119; CATATPASE.
DR PROSITE; PS00154; ATPASE_E1_E2; UNKNOWN_1.
KW Complete proteome.
SQ SEQUENCE 945 AA; 102476 MW; 2DBDC59DF35B9C9B CRC64;

Query Match 37.38; Score 44; DB 2; Length 945;
Best Local Similarity 58.88; Pred. No. 78;
Matches 10; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 5 SCEQQAENVQYHKNEL 21
    :|||:|||||
Db 46 SQQQAENVQYHKNEL 62

RESULT 34
Q9FD23 PRELIMINARY; PRT; 1057 AA.
AC Q9FD23;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DE 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE PUTATIVE HISTIDINE KINASE RECEPTOR.
GN WOL.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

```

```

OC eucosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE-ROOT;
RX PubMed=111114883;
RA Mahonen A., Bonke M.A., Kauppinen L., Riikonen M., Benfey P.N.,
RA Helariutta Y.;
RT "A novel two component hybrid molecule regulates the vascular pattern
RT of the Arabidopsis root.";
RL Genes Dev. 14:2938-2943(2000).
CC -1- SIMILARITY: THE N-TERMINAL REGION IS SIMILAR TO THAT OF OTHER
CC REGULATORY COMPONENTS OF SENSORY TRANSDUCTION SYSTEMS.
DR EMBL; AJ278530; CAC18523.1; -
DR EMBL; AJ278528; CAC18521.1; -
DR EMBL; AJ278529; CAC18522.1; -
DR InterPro; IPR000410; Bctrl_sens.
DR InterPro; IPR003594; HATPase_c.
DR InterPro; IPR003661; His_kinA.
DR InterPro; IPR001789; Response_reg.
DR Pfam; PF00072; response_reg; 1.
DR PRINTS; PR00344; BCTRLSENSOR.
DR SMART; SM00387; HATPase_c; 1.
DR SMART; SM00388; HSKA; 1.
DR SMART; SM00448; REC; 1.
DR Kinase; Phosphorylation; Receptor; Sensory transduction.
KW Kinase; Phosphorylation; Receptor; Sensory transduction.
SQ SEQUENCE 1057 AA; 117954 MW; E1381C8685F0D787 CRC64;

Query Match 37.38; Score 44; DB 10; Length 1057;
Best Local Similarity 29.48; Pred. No. 87;
Matches 10; Conservative 3; Mismatches 1; Indels 20; Gaps 1;

QY 6 CEQQAENVQYHKNEL 19
    :|||:|||||
Db 142 CDQARMQLQDFSVSVNHHALAILVSTFHYHKN 175

RESULT 35
Q9C5T9 PRELIMINARY; PRT; 1057 AA.
AC Q9C5T9;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE CYTOKININ RECEPTOR CREIA.
GN CREL.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC eucosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=CV. WASSILEWSKIIA;
RA Inoue T., Higuchi M., Hashimoto Y., Seki M., Kobayashi M., Kato T.,
RA Satoh T., Shinozaki K., Kakimoto T.;
RT "Identification of a cytokinin receptor, CREL, from Arabidopsis.";
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: THE N-TERMINAL REGION IS SIMILAR TO THAT OF OTHER
CC REGULATORY COMPONENTS OF SENSORY TRANSDUCTION SYSTEMS.
DR EMBL; AB049934; BAB33310.1; -
DR InterPro; IPR000410; Bctrl_sens.
DR InterPro; IPR003594; HATPase_c.
DR InterPro; IPR003661; His_kinA.
DR InterPro; IPR001789; Response_reg.
DR Pfam; PF02518; HATPase_c; 1.
DR Pfam; PF00072; response_reg; 1.
DR PRINTS; PR00344; BCTRLSENSOR.
DR SMART; SM00387; HATPase_c; 1.

```

```

DR SMART; SM00388; HSKA; 1.
DR SMART; SM00448; REC; 1.
SQ Phosphorylation; Receptor; Sensory transduction.
KW SEQUENCE 1057 AA; 117894 MW; E121A86345F0C292 CRC64;

Query Match      37.3%; Score 44; DB 10; Length 1057;
Best Local Similarity 29.4%; Pred. No. 87;
Matches 10; Conservative 3; Mismatches 1; Indels 20; Gaps 1;

QY 6 CEQQRNVQ-----HYHKN 19
   |:|:|:|:|
Db 142 CDQARMQLQDFSVSVNHHALAILVSTFHYHKN 175

RESULT 36
ID Q9C5U0 PRELIMINARY; PRT; 1080 AA.
AC Q9C5U0;
DT 01-JUN-2001 (TRENBLrel. 17, Created)
DT 01-JUN-2001 (TRENBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TRENBLrel. 17, Last annotation update)
DE HISTIDINE KINASE.
GN AHK4.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
ON NCBI_TaxID=3702;
RN SENSE FROM N.A.
RP SEQUENCE FROM N.A.
RC STRAIN=CV. WASSILEWSKIIJA;
RA Inoue T., Higuchi M., Hashimoto Y., Seki M., Kobayashi M., Kato T.,
RA Sato T., Shinzaki K., Kakimoto T.;
RT Identification of a cytokinin receptor, CRE1, from Arabidopsis.;
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
CC -|- SIMILARITY: THE N-TERMINAL REGION IS SIMILAR TO THAT OF OTHER
CC REGULATORY COMPONENTS OF SENSORY TRANSDUCTION SYSTEMS.
DR EMBL; AB049935; BAB3311.1;
DR InterPro; IPR000410; Bctrl_sens.
DR InterPro; IPR003594; HATPase_c.
DR InterPro; IPR003661; His_kinA.
DR InterPro; IPR001789; Response_reg.
DR Pfam; PF02518; HATPase_c; 1.
DR Pfam; PF00072; response_reg; 1.
DR PRINTS; PR00344; BCTRLSENSOR.
DR SMART; SM00387; HATPase_c; 1.
DR SMART; SM00388; HSKA; 1.
DR SMART; SM00448; REC; 1.
KW Phosphorylation; Receptor; Sensory transduction.
SQ SEQUENCE 1080 AA; 120670 MW; 59496F734B528114 CRC64;

Query Match      37.3%; Score 44; DB 10; Length 1080;
Best Local Similarity 29.4%; Pred. No. 89;
Matches 10; Conservative 3; Mismatches 1; Indels 20; Gaps 1;

QY 6 CEQQRNVQ-----HYHKN 19
   |:|:|:|:|
Db 165 CDQARMQLQDFSVSVNHHALAILVSTFHYHKN 198

RESULT 38
ID Q93787 PRELIMINARY; PRT; 139 AA.
AC Q93787;
DT 01-NOV-1996 (TRENBLrel. 01, Created)
DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TRENBLrel. 17, Last annotation update)
DE 2S ALBUMIN STORAGE PROTEIN PRECURSOR.
GN MAT5-D.
OS Gossypium hirsutum (Upland cotton).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Malvales; Malvaceae; Gossypium.
ON NCBI_TaxID=3635;
RN SENSE FROM N.A.
RP SEQUENCE FROM N.A.
RA Galau G.A., Wang H.Y.C., Hughes D.W.;
RL Submitted (JAN-1992) to the EMBL/GenBank/DBJ databases.
DR EMBL; M83301; AAA33049.1;
DR Mendel; 14688; Goshi; 2374; 14688.
DR InterPro; IPR001768; Cereal_tryp_amyl_inh.
DR InterPro; IPR003612; AAI.
DR Pfam; PF00234; tryp_alpha_amyl; 1.
DR SMART; SM00499; AAI; 1.
DR Signal.
FT SIGNAL.
FT CHAIN 1 20 POTENTIAL.
FT CHAIN 33 59 2S ALBUMIN STORAGE PROTEIN.
FT CHAIN 64 139 2S ALBUMIN STORAGE PROTEIN.
SQ SEQUENCE 139 AA; 15831 MW; 43ACF35FE97D19B4 CRC64;

Query Match      36.4%; Score 43; DB 10; Length 139;
Best Local Similarity 66.7%; Pred. No. 19;
Matches 8; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 4 DSCQQRNVQH 15

```

```

Db      34 DSCEQOIRKQH 45
||||| |
RESULT 39
O81055 PRELIMINARY; PRT; 170 AA.
AC O81055
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-JUN-2001 (TREMBLrel. 08, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE KINETCHORE (SKP1P)-LIKE PROTEIN.
GN T18E12.14
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Rounsley S.D., Lin X., Kaul S., Shea T.P., Fujii C.Y., Mason T.M.,
RA Shen M., Ronning C.M., Fraser C.M., Somerville C.R., Venter J.C.;
RT "Arabidopsis thaliana chromosome II BAC T18E12 genomic sequence.";
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC005313; AAC34483.1; -
DR Mendel; 31709; Arath; 1266; 31709.
DR InterPro; IPR001232; Skp1.
DR Pfam; PF01466; Skp1; 1.
DR SMART; SM00512; Skp1; 1.
SQ SEQUENCE 170 AA; 19398 MW; B75BEB9552CF7048 CRC64;

Query Match 36.4%; Score 43; DB 10; Length 170;
Best Local Similarity 47.6%; Pred. No. 23;
Matches 10; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

QY 1 NSPDSCQQQARNVQHYHKNEL 21
: || | | | |
Db 69 DSDDSTATSENVEAKNEL 89

RESULT 40
Q9D9J2 PRELIMINARY; PRT; 266 AA.
AC Q9D9J2
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE 1700061J05RIK PROTEIN.
GN 1700061J05RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=TESTIS;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino H., Adachi J., Fukuda S.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Kiyosawa H., Kondo S., Yamanaka I.,
RA Aizawa K., Izawa M., Nishi K., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Saito T., Okazaki Y., Ashburner M., Batalov S., Casavant T.,
RA Kadota K., Matsuda H.A., Shih C., King B., Kuchiwa H.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kuchiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
RA Schram L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bona M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,

```

```

RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL; AK006856; BAB24766.1; -
DR MGD; MGI:1920626; 1700061J05RIK.
SQ SEQUENCE 266 AA; 30227 MW; 72788236776163F0 CRC64;

Query Match 36.4%; Score 43; DB 11; Length 266;
Best Local Similarity 37.5%; Pred. No. 35;
Matches 6; Conservative 6; Mismatches 4; Indels 0; Gaps 0;

QY 2 SPDSCEQQQARNVQHYH 17
: | : | : | : |
Db 214 TPEVIEKSVRDVHWH 229

RESULT 41
O55014 PRELIMINARY; PRT; 319 AA.
AC O55014
DT 01-JUN-1998 (TREMBLrel. 06, Created)
DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE CARDIAC ANKYRIN REPEAT PROTEIN MCARP.
GN ALRP.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=HEART;
RX MEDLINE=97195688; PubMed=9043061;
RA Zou Y., Evans S., Chen J., Kuo H.C., Harvey R.P., Chien K.R.;
RT "CARP, a cardiac ankyrin repeat protein, is downstream in the Nkx2-5
RL homeobox gene pathway.";
RL Development 124:793-804(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=HEART;
RA Chen J., Chien K.R.;
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 103-176 FROM N.A.
RA Schoenfeld J.R., Lowe D.G., Zou Y., Chen J.;
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF041847; AAC03533.1; -
DR EMBL; AF041849; AAB97080.1; -
DR HSP; Q00421; 1AWC.
DR MGD; MGI:109621; ALRP.
DR InterPro; IPR002110; ANK.
DR Pfam; PF00023; ank; 4.
DR SMART; SM00248; ANK; 4.
DR PROSITE; PS50088; ANK_REPEAT; 4.
DR PROSITE; PS50297; ANK_REPEAT_REGION; 1.
SQ SEQUENCE 319 AA; 35939 MW; 6849DE07FD80517B CRC64;

Query Match 36.4%; Score 43; DB 11; Length 319;
Best Local Similarity 46.7%; Pred. No. 41;
Matches 7; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 1 NSPDSCQQQARNVQH 15
|||| | | | |
Db 144 NSPDVCDVEYKRTALH 158

RESULT 42
Q921F0

```

```

ID Q921FO PRELIMINARY; PRT; 319 AA.
AC Q921FO;
DT 01-MAY-1999 (TREMBlrel. 10, Created)
DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE CARP.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
ON NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY;
RA Jayaseelan R., Poizat C., Abdishoo S., Kedes L.;
RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; U50736; AAD10401.1; -
DR HSP; Q00421; IAWC.
DR InterPro; IPR002110; ANK.
DR Pfam; PF00023; ank; 4.
DR SMART; SM00248; ANK; 4.
DR PROSITE; PS00088; ANK_REPEAT; 4.
DR PROSITE; PS0297; ANK_REPEAT_REGION; 1.
SQ SEQUENCE 319 AA; 36075 MW; F78EB18BFD6049E0 CRC64;

Query Match 36.4%; Score 43; DB 11; Length 319;
Best Local Similarity 46.7%; Pred. No. 41;
Matches 7; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 1 NSPDSCEQQAQNVQH 15
|||||:::|
DB 144 NSPDVCDEYKRTALH 158

RESULT 43
Q9CR42 PRELIMINARY; PRT; 319 AA.
AC Q9CR42;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE ANKYRN-LIKE REPEAT PROTEIN.
GN ALRP.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
ON NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSU=TONGUE;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaoka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikolaio I., Pesole G., Quackenbush J.,
RA Schriml L.M., Stauber F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldairelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-P.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA Hayashizaki Y.;
RT Functional annotation of a full-length mouse cDNA collection."
RL ENBL; AK009959; BAB26611.1; -
DR EMBL; AK009655; BAB26419.1; -

```

```

MGD; MGI:109621; Alrp.
DR InterPro; IPR002110; ANK.
DR Pfam; PF00023; ank; 4.
DR SMART; SM00248; ANK; 4.
DR PROSITE; PS00088; ANK_REPEAT; 4.
DR PROSITE; PS0297; ANK_REPEAT_REGION; 1.
SQ SEQUENCE 319 AA; 36003 MW; 26A3C4062CF0E7D0 CRC64;

Query Match 36.4%; Score 43; DB 11; Length 319;
Best Local Similarity 46.7%; Pred. No. 41;
Matches 7; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 1 NSPDSCEQQAQNVQH 15
|||||:::|
DB 144 NSPDVCDEYKRTALH 158

RESULT 44
Q9XC26 PRELIMINARY; PRT; 326 AA.
AC Q9XC26;
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE AREA.
GN AREA.
OS Acinetobacter sp. ADP1.
OC Bacteria; Proteobacteria; gamma subdivision; Moraxellaceae;
OC Acinetobacter.
OX NCBI_TaxID=62977;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ADP1;
RA Jones R.M., Collier L.S., Neidle E.L., Williams P.A.;
RT "areABC genes determine the catabolism of aryl esters in Acinetobacter
RT sp. ADP1.";
RL J. Bacteriol. 0:0-0(1999).
DR EMBL; AF150928; AAD34027.1; -
DR InterPro; IPR000379; Est_lip_thioest_actsite.
DR InterPro; IPR002168; Lipolytic_enzyme.
DR PROSITE; PS01174; LIPASE_GDXG_SER; UNKNOWN1.
DR PROSITE; PS01174; LIPASE_GDXG_SER; UNKNOWN1.
SQ SEQUENCE 326 AA; 37066 MW; FF00B3549864AA57 CRC64;

Query Match 36.4%; Score 43; DB 2; Length 326;
Best Local Similarity 50.0%; Pred. No. 42;
Matches 9; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 3 PDSCEQQAQNVQH 20
|||||:::|
DB 219 PLUSAQMQRKIQLKNE 236

RESULT 45
Q94244 PRELIMINARY; PRT; 327 AA.
AC Q94244;
DT 01-FEB-1997 (TREMBlrel. 02, Created)
DT 01-FEB-1997 (TREMBlrel. 02, Last sequence update)
DT 01-NOV-1998 (TREMBlrel. 08, Last annotation update)
DE COSMID F57F4.
GN F57F4.2.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Alnscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsley T., Cooper J., Coulson A.,

```


RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
 RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
 RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
 RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
 RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,
 RA Smaldon N., Smith A., Sonhammer E., Staden R., Sulston J.,
 RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
 RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.,
 RT *2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 RT elegans.*;
 RL Nature 368:32-38(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA Miller N., Bradshaw H.;
 RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA Waterston R.;
 RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.
 DR EMBL; U70856; AAB09165.1; -;
 SQ SEQUENCE 327 AA; 36240 MW; C40E205FB4DC8B67 CRC64;

Query Match 36.4%; Score 43; DB 5; Length 327;
 Best Local Similarity 58.3%; Pred. No. 42;
 Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 8 QOARNVOHYHKN 19
 : : : : :
 Db 201 RKAASVAHYHKN 212

RESULT 46
 ID O51587 PRELIMINARY; PRT; 347 AA.
 AC O51587;
 DT 01-JUN-1998 (TREMBLrel. 06, Created)
 DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE SPERMIDINE/PUTRESCINE ABC TRANSPORTER, ATP-BINDING PROTEIN (POTA).
 GN BB0642
 OS Borrelia burgdorferi (Lyme disease spirochete).
 OC Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.
 OX NCBI_TaxID=139;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-ATCC 35210 / B31;
 RX MEDLINE-98065943; PubMed-9403685;
 RA Fraser C.M., Casjens S., Huang W.M., Sutton G.G., Clayton R.A.,
 RA Lathigra R., White O., Ketchum K.A., Dodson R., Hickey E.K., Gwinn M.,
 RA Dougherty B., Tomb J.-F., Fleischmann R.D., Richardson D.,
 RA Peterson J., Kerlavage A.R., Quackenbush J., Salzberg S., Hanson M.,
 RA van Vugt R., Palmer N., Adams M.D., Gocayne J.D., Weidman J.,
 RA Uterback T., Matthey L., McDonald L., Artiach P., Bowman C.,
 RA Garland S., Fujii C., Cotton M.D., Horst K., Roberts K., Hatch B.,
 RA Smith H.O., Venter J.C.;
 RT "Genomic sequence of a Lyme disease spirochaete, Borrelia
 burgdorferi".
 RL Nature 390:580-586(1997).
 CC -1- SIMILARITY: TO THE ATP-BINDING TRANSPORT PROTEIN FAMILY (ABC
 CC TRANSPORTERS).
 DR EMBL; AE001165; AAB91525.1; -;
 DR TIGR; BB0642; -;
 DR InterPro; IPR003593; AAA.
 DR InterPro; IPR003439; ABC-transporter.
 DR Pfam; PF00005; ABC_tran; 1.
 DR SMART; SM00382; AAA; 1.
 DR PROSITE; PS00211; ABC_TRANSPORTER; 1.
 KW ATP-binding; Complete proteome; Transport.
 SQ SEQUENCE 347 AA; 39580 MW; 5EA65C9F8BC04DAA CRC64;

Query Match 36.4%; Score 43; DB 2; Length 347;
 Best Local Similarity 37.5%; Pred. No. 44;
 Matches 6; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

QY 4 DSCEQOQARNVOHYHKN 19
 : : : : :
 Db 2 DNCILEIKNLSHYDN 17

RESULT 47
 ID Q9XZX7 PRELIMINARY; PRT; 353 AA.
 AC Q9XZX7;
 DT 01-NOV-1999 (TREMBLrel. 12, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
 DE HYPOTHETICAL 39.5 KDA PROTEIN.
 GN L2743.08
 OS Leishmania major.
 OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
 OX NCBI_TaxID=5664;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-FRIEDLIN;
 RA Ivens A.C., Murphy L., Quail M., Harris D., Oliver K., Lawson D.,
 RA Rajandream M.A., Barrell B.G.;
 RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-FRIEDLIN;
 RX MEDLINE-98146435; PubMed-9477341;
 RA Ivens A.C., Lewis S.M., Bagherzadeh A., Zhang L., Chan H.M.,
 RA Smith D.F.;
 RT "A physical map of the Leishmania major Friedlin genome.";
 RL Genome Res. 8:135-145(1998).
 DR EMBL; AL389894; CAC22689.1; -;
 KW Hypothetical protein.
 SQ SEQUENCE 353 AA; 39520 MW; 07D1D1961DE69A77 CRC64;

Query Match 36.4%; Score 43; DB 5; Length 353;
 Best Local Similarity 47.1%; Pred. No. 45;
 Matches 8; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 4 DSCEQOQARNVOHYHKN 20
 : : : : :
 Db 93 DEVNOLANNVEHPKSE 109

RESULT 48
 ID Q65153 PRELIMINARY; PRT; 448 AA.
 AC Q65153;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)
 DE PM448R.
 GN M448R.
 OS African swine fever virus (ASFV).
 OC Viruses; dsDNA viruses, no RNA stage; Asfarviridae;
 OC African swine fever-like viruses.
 OX NCBI_TaxID=10497;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BA71V;
 RX MEDLINE-96036500; PubMed-7483270;
 RA Yanez R.J., Rodriguez J.M., Nogal M.L., Yuste L., Enriquez C.,
 RA Rodriguez J.F., Vinuela E.;
 RT "Immune protection conferred by the baculovirus-related glycoprotein
 of Thogoto virus (Orthomyxoviridae).";
 RL Virology 208:249-278(1995).

RN RP SEQUENCE FROM N.A.
RC STRAIN-BA71V;
RX MEDLINE-94233765; PubMed-8178480;
RA La Vega I., Gonzalez A., Blasco R., Calvo V., Vinuela E.;
RT "Nucleotide sequence and variability of the inverted terminal
RL repetitions of African swine fever virus DNA.";
RN Virology 201:152-156(1994).
[3]
RN SEQUENCE FROM N.A.
RP STRAIN-BA71V;
RC MEDLINE-90219205; PubMed-2325203;
RA Gonzalez A., Calvo V., Almazan F., Almendral J.M., Ramirez J.C.,
RA La Vega I., Blasco R., Vinuela E.;
RT "Multigene families in African swine fever virus: family 360.";
RL J. Virol. 64:2073-2081(1990).
[4]
RN SEQUENCE FROM N.A.
RP STRAIN-BA71V;
RC MEDLINE-90219204; PubMed-2325202;
RA Almendral J.M., Almazan F., Blasco R., Vinuela E.;
RT "Multigene families in African swine fever virus: family 110.";
RL J. Virol. 64:2064-2072(1990).
[5]
RN SEQUENCE FROM N.A.
RP STRAIN-BA71V;
RC MEDLINE-91134988; PubMed-1994575;
RA Camacho A., Vinuela E.;
RT "Protein p22 of African swine fever virus: an early structural protein
RL that is incorporated into the membrane of infected cells.";
RN Virology 181:251-257(1991).
[6]
RN SEQUENCE FROM N.A.
RP STRAIN-BA71V;
RC MEDLINE-94187118; PubMed-8139051;
RA Rodriguez J.M., Yanez R.J., Pan R., Rodriguez J.F., Salas M.L.,
RA Vinuela E.;
RT "Multigene families in African swine fever virus: family 505.";
RL J. Virol. 68:2746-2751(1994).
[8]
RN SEQUENCE FROM N.A.
RP STRAIN-BA71V;
RC MEDLINE-93346971; PubMed-8393914;
RA Yanez R.J., Rodriguez J.M., Rodriguez J.F., Salas M.L., Vinuela E.;
RT "African swine fever virus thymidylate kinase gene: sequence and
RL transcriptional mapping.";
RN J. Gen. Virol. 74:1633-1638(1993).
[9]
RN SEQUENCE FROM N.A.
RP STRAIN-BA71V;
RC MEDLINE-94065656; PubMed-8245848;
RA Alcamí A., Angulo A., Vinuela E.;
RT "Mapping and sequence of the gene encoding the African swine fever
RL virion protein of M(r) 11500.";
RN J. Gen. Virol. 74:2317-2324(1993).
[10]
RN SEQUENCE FROM N.A.
RP STRAIN-BA71V;
RC MEDLINE-93277388; PubMed-8503790;
RA Munoz M., Freije J.M., Salas M.L., Vinuela E., Lopez-Otin C.;
RT "Structure and expression in E. coli of the gene coding for protein
RL p10 of African swine fever virus.";
RN Arch. Virol. 130:93-107(1993).
[11]
RN SEQUENCE FROM N.A.
RP STRAIN-BA71V;
RC MEDLINE-90357780; PubMed-2389555;
RA Blasco R., Lopez-Otin C., Munoz M., Bockamp E.O., Simon-Mateo C.,

RN RA Vinuela E.;
RT "Sequence and evolutionary relationships of African swine fever virus
RL thymidine kinase.";
RN Virology 178:301-304(1990).
[12]
RN SEQUENCE FROM N.A.
RP STRAIN-BA71V;
RC MEDLINE-93281390; PubMed-8506138;
RA Yanez R.J., Boursnell M., Nogal M.L., Yuste L., Vinuela E.;
RT "African swine fever virus encodes two genes which share significant
RL homology with the two largest subunits of DNA-dependent RNA
RN polymerases.";
RN Nucleic Acids Res. 21:2423-2427(1993).
[13]
RN SEQUENCE FROM N.A.
RP STRAIN-BA71V;
RC MEDLINE-93353606; PubMed-8102411;
RA Rodriguez J.M., Yanez R.J., Almazan F., Vinuela E., Rodriguez J.F.;
RT "African swine fever virus encodes a CD2 homolog responsible for the
RL adhesion of erythrocytes to infected cells.";
RN J. Virol. 67:5312-5320(1993).
[14]
RN SEQUENCE FROM N.A.
RP STRAIN-BA71V;
RC MEDLINE-94085774; PubMed-8262374;
RA Yanez R.J., Rodriguez J.M., Boursnell M., Rodriguez J.F., Vinuela E.;
RT "Two putative African swine fever virus helicases similar to yeast
RL 'DEAH' pre-mRNA processing proteins and vaccinia virus ATPases D1L
RN Gene 134:161-174(1993).
[15]
RN SEQUENCE FROM N.A.
RP STRAIN-BA71V;
RC MEDLINE-90223993; PubMed-2327074;
RA Lopez-Otin C., Freije J.M., Parra F., Mendez E., Vinuela E.;
RT "Mapping and sequence of the gene coding for protein p72, the major
RL capsid protein of African swine fever virus.";
RN Virology 175:477-484(1990).
[16]
RN SEQUENCE FROM N.A.
RP STRAIN-BA71V;
RC MEDLINE-94123986; PubMed-8293992;
RA Rodriguez J.M., Yanez R.J., Rodriguez J.F., Vinuela E., Salas M.L.;
RT "The DNA polymerase-encoding gene of African swine fever virus:
RL sequence and transcriptional analysis.";
RN Gene 136:103-110(1993).
[17]
RN SEQUENCE FROM N.A.
RP STRAIN-BA71V;
RC MEDLINE-93327788; PubMed-8335009;
RA Simon-Mateo C., Andres G., Vinuela E.;
RT "Polyprotein processing in African swine fever virus: a novel gene
RL expression strategy for a DNA virus.";
RN EMBO J. 12:2977-2987(1993).
[18]
RN SEQUENCE FROM N.A.
RP STRAIN-BA71V;
RC MEDLINE-93233210; PubMed-8474154;
RA Prados F.J., Vinuela E., Alcamí A.;
RT "Sequence and characterization of the major early phosphoprotein p32
RL of African swine fever virus.";
RN J. Virol. 67:2475-2485(1993).
[19]
RN SEQUENCE FROM N.A.
RP STRAIN-BA71V;
RC MEDLINE-92260660; PubMed-1583732;
RA Alcamí A., Angulo A., Lopez-Otin C., Munoz M., Freije J.M.,
RA Carrascosa A.L., Vinuela E.;
RT "Amino acid sequence and structural properties of protein p12, an
RL African swine fever virus attachment protein.";
RN J. Virol. 66:3860-3868(1992).
[20]
RN SEQUENCE FROM N.A.

RC	STRAIN-BA71V;
EX	MEDLINE-93174976; PubMed=8438592;
RA	Ianez R.J., Vinuela E.;
RT	"African swine fever virus encodes a DNA ligase.";
RL	Virology 193:531-536(1993).
RN	[21]
RP	SEQUENCE FROM N.A.
RC	STRAIN-BA71V;
RX	MEDLINE-93174941; PubMed=8382399;
RA	Pena L., Ianez R.J., Revilla Y., Vinuela E., Salas M.L.;
RT	"African swine fever virus guanylyltransferase.";
RL	Virology 193:319-328(1993).
RN	[22]
RP	SEQUENCE FROM N.A.
RC	STRAIN-BA71V;
RX	MEDLINE-95159428; PubMed=7856088;
RA	Simon-Mateo C., Freije J.M., Andres G., Lopez-Otin C., Vinuela E.;
RT	"Mapping and sequence of the gene encoding protein p17, a major
RL	African swine fever virus structural protein.";
RN	Virology 206:1140-1144(1995).
RN	[23]
RP	SEQUENCE FROM N.A.
RC	STRAIN-BA71V;
EX	MEDLINE-92263807; PubMed=1316688;
RA	Garcia-Beato R., Freije J.M., Lopez-Otin C., Blasco R., Vinuela E.,
RA	Salas M.L.;
RT	"A gene homologous to topoisomerase II in African swine fever virus.";
RL	Virology 188:938-947(1992).
RN	[24]
RP	SEQUENCE FROM N.A.
RC	STRAIN-BA71V;
EX	MEDLINE-94091056; PubMed=8256720;
RA	Freije J.M., Lain S., Vinuela E., Lopez-Otin C.;
RT	"Nucleotide sequence of a nucleoside triphosphate phosphohydrolase
RL	sequence from N.A."
RN	[25]
RP	Query Match 36.4%; Score 43; DB 12; Length 448;
Best Local Similarity 41.2%; Pred.No. 56;	
Matches 7; Conservative 4; Mismatches 6; Indels 0; Gaps 0;	
OY	3 PDSCEQOARNVQHYHKN 19
DB	116 PAPFOAQATROVQHFN 132
RESULT 49	
ID	Q9Z5G6 PRELIMINARY; PRT; 452 AA.
AC	Q9Z5G6;
DT	01-MAY-1999 (TEMBLrel. 10, Created)
DT	01-MAY-1999 (TEMBLrel. 10, Last sequence update)
DT	01-JUN-2001 (TEMBLrel. 17, Last annotation update)
DE	POTATIVE SERINE PROTEASE.
GN	MICB373.28.
OS	Mycobacterium leprae.
OC	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC	Actinomycetales; Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
OX	NCBI_TaxID=1769;
RA	Harris D., Taylor K.;
RP	Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
RN	[1]
RP	SEQUENCE FROM N.A.
RA	James K.D., Parkhill J., Barrell B.G., Rajandream M.A.;
RL	Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
RN	[3]
RP	SEQUENCE FROM N.A.
RA	MEDLINE-93188700; PubMed=8446027;
RX	Eigmeier K., Honore N., Woods S.A., Caudron B., Cole S.T.;
RT	"Use of an ordered cosmid library to deduce the genomic organization
RL	of Mycobacterium leprae.";
CL	Mol. Microbiol. 7:197-206(1993).
-I-	SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY.

DR	EMBL; AL035500; CAB36690.1; -
DR	InterPro; IPR001478; PDZ.
DR	InterPro; IPR001940; Protease2C.
DR	InterPro; IPR001254; Trypsin.
DR	Pfam; PF00595; PDZ; 1
DR	Pfam; PF00089; PDZ; 1
DR	PRINTS; PR00834; trypsin; 1
DR	SMART; SM00228; PDZ; 1.
KW	Hydrolase; Protease; Serine protease.
SQ	SEQUENCE 452 AA; 45129 MW; 6CA675EB0911F983 CRC64;
Query Match	36.4%; Score 43; DB 2; Length 452;
Best Local Similarity	46.7%; Pred.No. 57;
Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps	
OY	4 DSCEQOARNVQHYHK 18
DB	41 DWCYQSPSLQOYHR 55
RESULT 50	
ID	Q9V5J4 PRELIMINARY; PRT; 743 AA.
AC	Q9V5J4;
DT	01-MAY-2000 (TEMBLrel. 13, Created)
DT	01-MAY-2000 (TEMBLrel. 13, Last sequence update)
DT	01-MAR-2001 (TEMBLrel. 16, Last annotation update)
DE	CG3298 PROTEIN.
GN	JHI-1 OR CG3298.
OS	Drosophila melanogaster (Fruit fly).
OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC	Ephydroidea; Drosophilidae; Drosophila.
OX	NCBI_TaxID=7227;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN-BERKELEY;
RX	MEDLINE-20196006; PubMed=10731132;
RA	Adams M.D., Celnikner S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA	Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA	George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA	Sutton G.G., Wortman J.R., Yeaman M.D., Zhang Q., Chen L.X.,
RA	Brandon R.C., Rogers Y.H.C., Blazej R.G., Champagne M., Pfeiffer B.D.,
RA	Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA	Abriel J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA	Balwell R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA	Beeson K.Y., Benos P.V., Bereman B.P., Bhandari D., Bolshakov S.,
RA	Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA	Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA	Cherry J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,
RA	de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA	Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA	Dubin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA	Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA	Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA	Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA	Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA	Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA	Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA	Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA	Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA	Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA	Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA	Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA	Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA	Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA	Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA	Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA	Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA	Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA	Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA	

107 37 31.4 540 1 HXTD_YEAST
 108 37 31.4 554 1 Y514_SYNY3
 109 37 31.4 570 1 RM44_CAEEL
 110 37 31.4 594 1 YBJK_DROME
 111 37 31.4 623 1 DRTS_PLAVI
 112 37 31.4 633 1 IPAA_SHIEL
 113 37 31.4 640 1 HS74_ANOAL
 114 37 31.4 642 1 YB65_SCHPO
 115 37 31.4 647 1 ZG48_XENLA
 116 37 31.4 761 1 CFAB_MOUSE
 117 37 31.4 779 1 ZW10_HUMAN
 118 37 31.4 805 1 E2F_DROME
 119 37 31.4 864 1 LDLR_MOUSE
 120 37 31.4 1039 1 Y304_TREPA
 121 37 31.4 1043 1 TCF8_MESAU
 122 37 31.4 1109 1 TCF8_RAT
 123 37 31.4 1114 1 TCF8_CHICK
 124 37 31.4 1231 1 YKT3_CAEEL
 125 37 31.4 1493 1 M3K1_RAT
 126 37 31.4 1495 1 M3K1_HUMAN
 127 37 31.4 2331 1 RRPL_MABVM
 128 37 31.4 2331 1 RRPL_MABVP
 129 37 31.4 3056 1 ATM_HUMAN
 130 37 31.4 5327 1 AC7_MOUSE
 131 37 31.4 5430 1 AC7_HUMAN
 132 36.5 30.9 106 1 GAS4_ARATH
 133 36.5 30.9 143 1 Y082_RICPR
 134 36.5 30.9 235 1 YDE2_SCHPO
 135 36.5 30.9 291 1 SIX2_HUMAN
 136 36.5 30.9 296 1 SIX2_MOUSE
 137 36.5 30.9 299 1 VV_RINDR
 138 36.5 30.9 310 1 HEMZ_BACSU
 139 36.5 30.9 507 1 RRPP_RINDK
 140 36.5 30.9 507 1 RRPP_RINDR
 141 36.5 30.9 689 1 PPCE_AERHY
 142 36.5 30.9 867 1 VL96_IRV1
 143 36.5 30.9 1059 1 CERU_RAT
 144 36.5 30.9 1137 1 A4E1_HUMAN
 145 36 30.5 96 1 SY20_RAT
 146 36 30.5 112 1 Y13K_SSV1
 147 36 30.5 157 1 YQFG_BACSU
 148 36 30.5 167 1 LKCB_PASHA
 149 36 30.5 175 1 YBP7_YEAST
 150 36 30.5 189 1 GRPE_HELPY

ALIGNMENTS

RESULT 1
 TBX8_CAEEL
 ID TBX8_CAEEL STANDARD; PRT; 342 AA.
 AC Q22292;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE T-BOX PROTEIN 8
 GN TBX-8 OR T07C4.2
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA Berks M.;
 RL Submitted (JAN-1994) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP REVISIONS.
 RC STRAIN-BRISTOL N2;
 RA Durbin R.;
 RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).

CC -1- SIMILARITY: CONTAINS A T-BOX DOMAIN.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC EMBL; Z29443; CAA82578.1; -
 DR HSSP; P24781; 1XBR.
 DR WormPep; T07C4.2; CE21157.
 DR InterPro; IPR001699; T-box.
 DR Pfam; PF00907; T-box; 1.
 DR PRINTS; PR00937; TBOX.
 DR SMART; SM00425; TBOX; 1.
 DR PROSITE; PS01283; TBOX_1; 1.
 DR PROSITE; PS01264; TBOX_2; 1.
 DR PROSITE; PS02522; TBOX_3; 1.
 DR DNA-binding; Nuclear protein.
 KW DNA_BIND 16 195
 FT DNA_BIND 16 195
 FT SEQUENCE 342 AA; 38923 MW; 926DA46941277C49 CRC64;
 SQ
 Query Match 38.1%; Score 45; DB 1; Length 342;
 Best Local Similarity 70.0%; Pred. No. 8.9; 1; Indels 0; Gaps 0;
 Matches 7; Conservative 2; Mismatches 1;
 QY 12 NVQHYHKNEL 21
 I: |||||:
 Db 18 NLFHYHKNEM 27
 RESULT 2
 D7_DICDI
 ID D7_DICDI STANDARD; PRT; 850 AA.
 AC P54682;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE CAMP-INDUCIBLE PRESPORE PROTEIN D7 PRECURSOR.
 GN Dictyostelium discoideum (Slime mold).
 OS Eukaryota; Mycetozoa; Dictyosteliida; Dictyostellum.
 OX NCBI_TaxID=44689;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-AX3;
 RX MEDLINE=95080502; PubMed=7988791;
 RA Agarwal A., Sloger M.S., Oyama M., Blumberg D.D.;
 RT "Analysis of a novel cyclic Amp inducible prespore gene in
 RT Dictyostelium discoideum: evidence for different patterns of CAMP
 RT regulation.";
 RL Differentiation 57:151-162(1994).
 CC -1- DEVELOPMENTAL STAGE: EXPRESSED SPECIFICALLY IN THE PRESPORE CELLS.
 CC -1- INDUCTION: BY CAMP.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC EMBL; U25143; AAA73514.1; -
 DR DictyDb; DD02038; -
 KW Sporulation; Signal.
 FT SIGNAL 1 24
 FT CHAIN 25 850
 FT DOMAIN 470 475
 FT DOMAIN 555 568
 FT POTENTIAL.
 FT CAMP-INDUCIBLE PRESPORE PROTEIN D7.
 FT POLY-GLN.
 FT POLY-ASN.

FT DOMAIN 728 738 POLY-CLN.
SQ SEQUENCE 850 AA; 95343 MW; 13BA634CCET/AA502 CRC64;

Query Match 38.1%; Score 45; DB 1; Length 850;
Best Local Similarity 50.0%; Pred. No. 23;
Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 2 SPDSCEQOARNVQH 15
| | | : | : | : |
DB 810 SPDIKQLKNIQH 823

RESULT 3
DEFA_ANTMA STANDARD; PRT; 227 AA.
AC P23706;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE FLORAL HOMEOTIC PROTEIN DEFICIENS.
GN DEFA.
OS Antirrhinum majus (Garden snapdragon).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids I; Lamiales; Veroniceae; Antirrhinum.
OX NCBI_TaxID=4151;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90183955; PubMed=1968830;
RA Sommer H., Beltran J.-P., Huijser P., Pape H., Loennig W.-E.,
RA Saedler H., Schwarz-Sommer Z.;
RT "Deficiens, a homeotic gene involved in the control of flower
RT morphogenesis in Antirrhinum majus: the protein shows homology to
RT transcription factors.";
RL EMBO J. 9:605-613(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=SIPE50;
RX MEDLINE=92155166; PubMed=1346760;
RA Schwarz-Sommer Z., Hue I., Huijser P., Flor P.J., Hansen R.,
RA Tetens F., Loennig W.-E., Saedler H., Sommer H.;
RT "Characterization of the Antirrhinum floral homeotic MADS-box gene
RT deficiens: evidence for DNA binding and autoregulation of its
RL persistent expression throughout flower development.";
RL EMBO J. 11:231-263(1992).

CC -1- FUNCTION: TRANSCRIPTION FACTOR INVOLVED IN THE GENETIC CONTROL OF
CC FLOWER DEVELOPMENT. ACTS IN CONJUNCTION WITH GLOBOSA (GLO).
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- MISCELLANEOUS: MUTATIONS IN DEFA CAUSE TRANSFORMATION OF PETALS
CC INTO SEPALs AND STAMINA INTO CARPELS.
CC -1- SIMILARITY: BELONGS TO THE MADS DOMAIN FAMILY OF TRANSCRIPTION
CC FACTORS.
CC -1- SIMILARITY: CONTAINS A PROBABLE DIMERIZATION DOMAIN FOUND IN
CC SRF-TYPE TRANSCRIPTION FACTORS (K-BOX).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL; X52023; CAA36268.1; -
CC EMBL; X62810; CAA44629.1; -
CC PIR; S12378; S12378.
CC PIR; S19232; S19232.
CC HSSP; P11746; LMNM.
CC TRANSFAC; T01008; -
CC InterPro; IPR002487; K-box.
CC InterPro; IPR002100; MADS-box.
CC Pfam; PF01486; K-box; 1.

DR Pfam; PF00319; SRF-TF; 1.
DR PRINTS; PR00404; MADSDOMAIN.
DR SMART; SM00432; MADS; 1.
DR PROSITE; PS00350; MADS_BOX_1; 1.
DR PROSITE; PS0066; MADS_BOX_2; 1.
KW Transcription regulation; DNA-binding; Activator; Nuclear protein;
KW Developmental protein.
FT DOMAIN 3 57 MADS.
FT DOMAIN 93 165 K-BOX.
SQ SEQUENCE 227 AA; 26279 MW; 95E3FF60924FDE8D CRC64;

Query Match 37.3%; Score 44; DB 1; Length 227;
Best Local Similarity 36.8%; Pred. No. 8.3;
Matches 7; Conservative 6; Mismatches 6; Indels 0; Gaps 0;

QY 1 NSPDSCEQOARNVQHYKN 19
| | : : : | : | : |
DB 145 NQIDTSKKKVRNVEIHRN 163

RESULT 4
VG20_BPP22
ID VG20_BPP22 STANDARD; PRT; 471 AA.
AC Q01076;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE DNA TRANSFER PROTEIN GP20.
GN GP20.
OS Bacteriophage P22.
OC Viruses; dsDNA viruses, no RNA stage; Tailed phages; Podoviridae.
OX NCBI_TaxID=10754;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93219140; PubMed=8464750;
RA Adhikari P., Berget P.B.;
RT "Sequence of a DNA injection gene from Salmonella typhimurium phage
RT P22.";
RL Nucleic Acids Res. 21:1499-1499(1993).
RN [2]
RP SEQUENCE FROM N.A.
RA Kropinski A.M.B., VanderByl C.S.;
RT "The completed sequence of genome of Salmonella phage P22.";
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.

CC SEQUENCE OF 1-92 FROM N.A.
CC MEDLINE=92394890; PubMed=1522065;
CC Conlin C.A., Vimr E.R., Miller C.G.;
CC "Oligopeptidase A is required for normal phage P22 development.";
CC J. Bacteriol. 174:5869-5880(1992).
CC -1- FUNCTION: REQUIRED FOR EJECTION OF THE PHAGE DNA FROM THE PHAGE
CC PARTICLE AND INJECTION OF THAT DNA INTO THE HOST.
CC -1- SIMILARITY: STRONG, TO PHAGE APSE-1 P33.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL; L07556; AAA62407.1; -
CC EMBL; AF217253; AAF75054.1; -
CC EMBL; M93985; AAA72116.1; -
CC PIR; D43330; D43330.
CC Late protein.
SQ SEQUENCE 471 AA; 50100 MW; EA6366D094D10A9D CRC64;

Query Match 37.3%; Score 44; DB 1; Length 471;
Best Local Similarity 36.8%; Pred. No. 18;

Matches 7; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

QY 1 NSPDSCEQQARNVOHYHKN 19
:||||:| | | | |
Db 411 SSPEAMQOSIREIQEVTNN 429

```
RESULT 5
WNSC_XENLA STANDARD; PRT; 360 AA.
ID WNSC_XENLA
AC P33945; Q91928;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE WNT-5C PROTEIN PRECURSOR (XWNT-5C).
GN WNT-5C.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RA Koster J.G., Kuiken G.A., Stegeman B., Peterson J., Eizema K.,
RA Stabel L., Dekker E.J., Destre O.H.J.;
RA Submitted (JUN-1993) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 1-27 FROM N.A.
RX TISSUE=Embryo;
RC MEDLINE=94261437; PubMed=8202371;
RA Kuiken G.A., Bertens P.J.A., Peterson-Maduro J., Veenstra G.J.C.,
RA Koster J.G., Destree O.H.J.;
RT "The promoter of the Xwnt-5C gene contains octamer and AP-2 motifs
functional in Xenopus embryos."
RL Nucleic Acids Res. 22:1675-1680(1994).
CC -!- FUNCTION: PROBABLE DEVELOPMENTAL PROTEIN. MAY BE A SIGNALING
CC MOLECULE WHICH AFFECTS THE DEVELOPMENT OF DISCRETE REGIONS OF
CC TISSUES. IS LIKELY TO SIGNAL OVER ONLY FEW CELL DIAMETERS.
CC -!- SUBCELLULAR LOCATION: POSSIBLY SECRETED AND ASSOCIATES WITH THE
CC EXTRACELLULAR MATRIX.
CC -!- DEVELOPMENTAL STAGE: EXPRESSION IN THE EARLY GASTRULA STAGE
CC ONWARDS.
CC -!- SIMILARITY: BELONGS TO THE WNT FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X73510; CAA51916.1; -
CC EMBL; X76190; CAA53784.1; -
CC PIR; S34173.
CC InterPro; IPR000970; Wnt1.
CC Pfam; PF00110; wnt; 1.
CC SMART; SM00097; Wnt1; 1.
CC PROSITE; PS00246; WNT1; 1.
CC Developmental protein 1; 16
CC SIGNAL 17 360
CC CHAIN 17 360 WNT-5C PROTEIN.
CC FT CARBOHYD 94 94 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 100 100 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 292 292 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 306 306 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CONFLICT 15 15 S->C (IN REF. 2).
CC SQ SEQUENCE 360 AA; 40714 MW; 93CBD15D7A92779E CRC64;
```

Query Match 36.4%; Score 43; DB 1; Length 360;
Best Local Similarity 40.0%; Pred. No. 19;
Matches 8; Conservative 4; Mismatches 8; Indels 0; Gaps 0;

QY 1 NSPDSCEQQARNVOHYHKN 20
:||||:| | | | |
Db 185 NFPKSEEQARSLMNLQNE 204

```
RESULT 6
PUR8_HELPJ STANDARD; PRT; 440 AA.
ID PUR8_HELPJ
AC Q9ZKA2;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE ADENYLOSUCINATE LYASE (EC 4.3.2.2) (ADENYLOSUCINASE) (ASL).
GN PURB OR JHP1039.
OS Helicobacter pylori J99 (Campylobacter pylori J99).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_TaxID=85963;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=99120557; PubMed=9923682;
RA Alm R.A., Ling L.-S.L., Moir D.T., King B.L., Brown E.D., Doig P.C.,
RA Smith D.R., Noonan B., Guild B.C., deJonge B.L., Carmel G.,
RA Tummino P.J., Caruso A., Uria-Nickelsen M., Mills D.M., Ives C.,
RA Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Vovis G.F.,
RA Trust T.J.;
RA "Genomic sequence comparison of two unrelated isolates of the human
RT gastric pathogen Helicobacter pylori."
RL Nature 397:176-180(1999).
CC -!- CATALYTIC ACTIVITY: 1-(5'-PHOSPHORIBOSYL)-4-(N-SUCCINO-CARBOXAMIDE)
CC -5-AMINOIMIDAZOLE = FUMARATE + 5'-PHOSPHORIBOSYL-5-AMINO-4-
CC IMIDAZOLECARBOXAMIDE (ALSO CATALYZES: N6-(1,2-DICARBOXYETHYL)AMP =
CC FUMARATE + AMP).
CC -!- PATHWAY: EIGHT STEP IN DE NOVO PURINE BIOSYNTHESIS.
CC -!- SIMILARITY: BELONGS TO THE LYASE I FAMILY. ADENYLOSUCINATE LYASE
CC SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE001531; AAD06609.1; -
CC InterPro; IPR000362; Fumarate_lyase.
CC Pfam; PF00206; lyase_1; 1.
CC PRINTS; PR00145; DCRYSTALLIN.
CC PRINTS; PR00149; FUMRATLYASE.
CC PROSITE; PS00163; FUMARATE_LYASES; 1.
CC Purine biosynthesis; Lyase; Complete proteome.
CC FT ACT_SITE 68 ACID (BY SIMILARITY).
CC FT ACT_SITE 141 BASE (BY SIMILARITY).
CC SQ SEQUENCE 440 AA; 49867 MW; C67D3C02AEFA2EFE CRC64;
```

Query Match 36.4%; Score 43; DB 1; Length 440;
Best Local Similarity 41.2%; Pred. No. 24;
Matches 7; Conservative 6; Mismatches 4; Indels 0; Gaps 0;

QY 5 SCEQQARNVOHYHKNEL 21
:||||:| | | | |
Db 232 SCEKIANIRHLQRSEV 248

```
RESULT 7
PUR8_HELPJ STANDARD; PRT; 440 AA.
ID PUR8_HELPJ
AC P56468;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
```



```

DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE ADENYLOSUCCINATE LYASE (EC 4.3.2.2) (ADENYLOSUCCINASE) (ASL).
GN PURB OR HP1112.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_TaxID=210;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=26695 / ATCC 700392;
RX MEDLINE=97394467; PubMed=9252185;
RA Tomb J.-F., White O., Kerlavage A.R., Clayton R.A., Sutton G.G.,
RA Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,
RA Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,
RA Loftus B., Richardson D., Dodson R., Khalak H.G., Glodek A.,
RA McKenney K., Fitzgerald L.M., Lee N., Adams M.D., Hickey E.K.,
RA Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,
RA Cotton M.D., Weidman J.M., Fujii C., Bowman C., Watthey L., Wallin E.,
RA Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,
RA Venter J.C.;
RT "The complete genome sequence of the gastric pathogen Helicobacter
pylori."
RN [2]
RP SEQUENCE OF 1-21 FROM N.A.
RA Poppe M., Botz J., Hahm B., Dobat K., Eickelbaum W., Paweletz N.,
RA Arand M., Knehr M.;
RT "Promoter characterization of centromere protein C reveals its
participation in cell cycle regulation in late G1-phase and expression
control by E2F-1, pRB, p107 and Sp-1."
RL Submitted (MAY-1999) to the EMBL/Genbank/DBJ databases.
CC -!- FUNCTION: COMPONENT OF THE INNER KINETOCHORE PLATE. REQUIRED FOR
CC NORMAL KINETOCHORE ASSEMBLY.
CC -!- SUBCELLULAR LOCATION: NUCLEAR.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M95724; AAA51974.1;
DR EMBL; AF151723; AAF73191.1;
DR PIR; A42681; A42681.
DR MIM; 117141;
KW Chromosomal protein; Nuclear protein; DNA-binding; Phosphorylation;
KW Centromere.
FT DOMAIN 259 273 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT DOMAIN 484 499 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT DOMAIN 558 574 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT DOMAIN 780 798 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT MOD_RES 75 75 PHOSPHORYLATION (POTENTIAL).
FT MOD_RES 732 732 PHOSPHORYLATION (POTENTIAL).
SQ SEQUENCE 943 AA; 106925 MW; 6D80810A3E476376 CRC64;

Query Match 36.4%; Score 43; DB 1; Length 943;
Best Local Similarity 41.2%; Pred. No. 53;
Matches 7; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

QY 1 NSPDSCEQQARNVQHYH 17
   |::|::|::|::|
Db 129 NTPDSKKISSRNINDH 145

RESULT 9
YHJL_ECOLI
ID YHJL_ECOLI STANDARD; PRT; 1140 AA.
AC P37650; P76710;
DT 01-OCT-1994 (Rel. 30, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE HYPOTHETICAL 128.7 KDA PROTEIN IN DCTA-DPPF INTERGENIC REGION.
GN YHJL OR B3530.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=94316500; PubMed=8041620;
RA Sofia H.J., Burland V., Daniels D.L., Plunkett G. III, Blattner F.R.;
RT "Analysis of the Escherichia coli genome. V. DNA sequence of the
region from 76.0 to 81.5 minutes."

```

RL Nucleic Acids Res. 22:2576-2586(1994).

RN [2]

RP REVISIONS TO 577-578.

RC STRAIN-K12 / MG1655;

RA BLATTNER F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,

RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,

RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,

RA Mau B., Shao Y.;

RT "The complete genome sequence of *Escherichia coli* K-12."

RL Science 277:1453-1474(1997).

CC -!- SIMILARITY: STRONG, TO ACETOBACTER XYLINUM CELLULOSE SYNTHASE

CC OPERON PROTEIN C (ACSC/BCSC).

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL: U00039; CAB34649.1; ALT_INIT.

DR EMBL: AE000430; AAC7655.1; ALT_INIT.

DR EcoGene; EG12257; YhJL.

DR InterPro; IPR001440; TPR.

DR Pfam; PF00515; TPR: 3.

KW Hypothetical protein; Cellulose biosynthesis; Complete proteome.

FT CONFLICT 577 578 AM -> V (IN REF. 1).

CC SEQUENCE 1140 AA; 12581 MW; 6811A8B9F0AFE29B CRC64;

CC -----

Query Match 36.0%; Score 42.5; DB 1; Length 1140;

Best Local Similarity 55.6%; Pred. No. 77;

Matches 10; Conservative 2; Mismatches 3; Indels 3; Gaps 1;

CC

QY 1 NSPDSCE---QQARNVQH 15

DB 348 NNPDRERLFQARNVDN 365

CC

RESULT 10

ID YIEE_ECOLI STANDARD; PRT; 253 AA.

AC P31464; P76740;

DT 01-JUL-1993 (Rel. 26, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 20-AUG-2001 (Rel. 40, Last annotation update)

DE HYPOTHETICAL 28.2 KDA PROTEIN IN TNAB-BGLB INTERGENIC REGION.

OS *Escherichia coli*.

GN Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

OC Escherichia.

OX NCBI_TaxID=562;

CC

CC SEQUENCE FROM N.A.

CC STRAIN-K12 / MG1655;

CC MEDLINE-93315143; PubMed-7686882;

CC Burland V.D., Plunkett G. III, Daniels D.L., Blattner F.R.;

CC "DNA sequence and analysis of 136 kilobases of the *Escherichia coli*

CC genome: organizational symmetry around the origin of replication."

CC Genomics 16:551-561(1993).

CC [2]

CC REVISIONS TO 101-152.

CC STRAIN-K12 / MG1655;

CC MEDLINE-97426617; PubMed-9278503;

CC Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,

CC Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,

CC Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,

CC Mau B., Shao Y.;

CC "The complete genome sequence of *Escherichia coli* K-12."

CC Science 277:1453-1474(1997).

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL: L10328; AAA62063.1; --

DR EMBL: AE000448; AAC78735.1; --

DR EcoGene; EG11722; YieB.

KW Hypothetical protein; Complete proteome.

FT CONFLICT 101 152 GEGDLMELQRAIRGRHSPHAPDHTFSSNESLWISKNDP

FT FT FPMNRYGSGVNTILTKRGRHAA (IN REF. 1).

CC SEQUENCE 253 AA; 28191 MW; 5A4418BF7CD73CF6 CRC64;

CC -----

Query Match 35.6%; Score 42; DB 1; Length 253;

Best Local Similarity 35.3%; Pred. No. 19;

Matches 6; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

CC

QY 3 PDSCEQQARNVQHYHKN 19

DB 25 PSQCHQEARNIPPHRQS 41

CC

RESULT 11

ID WNSB_AMBME STANDARD; PRT; 357 AA.

AC Q06443;

DT 01-JUN-1994 (Rel. 29, Created)

DT 01-JUN-1994 (Rel. 29, Last sequence update)

DT 01-JUN-1994 (Rel. 29, Last annotation update)

DE WNT-5B PROTEIN PRECURSOR.

GN WNT-5B.

OS Ambystoma mexicanum (Axolotl).

CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

CC Amphibia; Batrachia; Caudata; Salamandroidea; Ambystomatidae;

CC Ambystoma.

OX NCBI_TaxID=8296;

CC

CC SEQUENCE FROM N.A.

CC MEDLINE-93183769; PubMed-8443107;

CC Busse U., Sequin C.;

CC "Isolation of cDNAs for two closely related members of the axolotl

CC Wnt family, Awnt-5A and Awnt-5B, and analysis of their expression

CC during development."

CC Mech. Dev. 40:63-72(1993).

CC -!- FUNCTION: PROBABLE DEVELOPMENTAL PROTEIN. MAY BE A SIGNALING

CC MOLECULE WHICH AFFECTS THE DEVELOPMENT OF DISCRETE REGIONS OF

CC TISSUES. IS LIKELY TO SIGNAL OVER ONLY FEW CELL DIAMETERS.

CC -!- SUBCELLULAR LOCATION: POSSIBLY SECRETED AND ASSOCIATES WITH THE

CC EXTRACELLULAR MATRIX.

CC -!- TISSUE SPECIFICITY: PREDOMINANTLY IN NEUROECTODERMAL TISSUES.

CC -!- DEVELOPMENTAL STAGE: UNDETECTABLE IN THE BLASTULA. APPEAR WITH

CC GASTRULATION, IS PRESENT THROUGHOUT NEURULATION AND ORGANOGENESIS,

CC AND DECREASE TO BARELY DETECTABLE LEVELS IN HATCHED LARVAE.

CC -!- SIMILARITY: BELONGS TO THE WNT FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL: Z14048; CAA78416.1; --

DR PIR; S25000; S25000.

DR InterPro; IPR000970; Wnt1.

DR Pfam; PF00110; wnt; 1.

DR SMART; SM00097; WNT1; 1.

CC

DR PROSITE; PS00246; WNT1; 1.
 KW Developmental protein; Glycoprotein; Signal.
 FT SIGNAL 1 18 POTENTIAL.
 FT CHAIN 19 357 WNT-5B PROTEIN.
 FT CARBOHYD 91 91 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 97 97 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 289 289 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 303 303 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 357 AA; 40086 MW; 3AAD6B8807BF7DB4 CRC64;

Query Match 35.6%; Score 42; DB 1; Length 357;
 Best Local Similarity 40.0%; Pred. No. 27;
 Matches 8; Conservative 3; Mismatches 9; Indels 0; Gaps 0;

OY 1 NSPDSCEQQAARVQVHYKNE 20
 | | | | | | | | | |
 DB 182 NYPKSEEQARTLMNLQNE 201

RESULT 12

VGLY_PIARV
 ID VGLY_PIARV STANDARD; PRT; 503 AA.
 AC P03540;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE GLYCOPROTEIN POLYPROTEIN PRECURSOR [CONTAINS: GLYCOPROTEINS G1 AND G2].
 DE GPC.
 GN Pichinde arenavirus.
 OS Viruses; ssRNA negative-strand viruses; Arenaviridae; Arenavirus.
 OC NCBI_TaxID=11630;
 OX [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=85033957; PubMed=6492264;
 RA Auperin D.D., Romanowski V., Galinski M., Bishop D.H.L.;
 RT *Sequencing studies of pichinde arenavirus S RNA indicate a novel
 RT coding strategy, an ambisense viral S RNA.*;
 RL J. Virol. 52:897-904(1984).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=87160943; PubMed=2435460;
 RA Bishop D.H.L., Auperin D.D.;
 RT *Arenavirus gene structure and organization.*;
 RL Curr. Top. Microbiol. Immunol. 133:5-17(1987).
 CC -I- SIMILARITY: BELONGS TO THE ARENAVIRUSES GPC PROTEIN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; K02734; AAA46824.1; .
 CC EMBL; M16735; AAA46827.1; .
 CC FIR; A04149; QQQPGP.
 DR InterPro; IPR001535; Arena_glycoprot.
 DR Pfam; PF00798; Arena_glycoprot; 1.
 KW Polypeptide; Glycoprotein; Envelope protein.
 FT CHAIN 1 271 GLYCOPROTEIN G1.
 FT CHAIN 272 503 GLYCOPROTEIN G2.
 FT CARBOHYD 67 67 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 74 74 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 89 89 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 100 100 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 111 111 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 116 116 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 121 121 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 132 132 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 181 181 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 217 217 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 241 241 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 379 379 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 387 387 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 404 404 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 409 409 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 499 499 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 503 AA; 57278 MW; 17740E092B450044 CRC64;

Query Match 35.6%; Score 42; DB 1; Length 503;
 Best Local Similarity 44.4%; Pred. No. 39;
 Matches 8; Conservative 2; Mismatches 6; Indels 2; Gaps 1;

OY 1 NSPDSCEQQAARVQVHYK 18
 | | | | | | | | | |
 DB 81 NLPQSCSK--NNTTHYK 96

RESULT 13

AC15_NEUCR
 ID AC15_NEUCR STANDARD; PRT; 865 AA.
 AC P87000;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE TRANSCRIPTIONAL ACTIVATOR PROTEIN ACU-15.
 GN ACU-15.
 OS Neurospora crassa.
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
 OC Sordariales; Sordariaceae; Neurospora.
 OX NCBI_TaxID=5141;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=74A / STA;
 RA Bibbins M., Connerton I.F.;
 RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
 CC -I- FUNCTION: POSITIVE REGULATOR OF ACETATE INDUCTION.
 CC -I- SUBCELLULAR LOCATION: NUCLEAR.
 CC -I- DOMAIN: THE GLUTAMINE-RICH DOMAIN MIGHT FUNCTION IN ACTIVATING
 CC GENE EXPRESSION.
 CC -I- SIMILARITY: CONTAINS 1 ZN(2)-CYS(6), FUNGAL-TYPE BINUCLEAR
 CC CLUSTER DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; Y11565; CAA72324.1; .
 CC HSSP; P04386; LAW6.
 CC InterPro; IPR001138; ZN2_CY6_fungal.
 CC Pfam; PF00172; Zn_clus; 1.
 CC PRINTS; PR00054; FUNGALZNCYS.
 CC SMART; SM00066; GAL4; 1.
 CC PROSITE; PS00463; ZN2_CY6_FUNGAL_1; 1.
 CC PROSITE; PS00448; ZN2_CY6_FUNGAL_2; 1.
 KW Transcription regulation; DNA-binding; Nuclear protein; Zinc;
 KW Metal-binding; Activator.
 FT DNA_BIND 24 51 ZN(2)-CYS(6), FUNGAL-TYPE.
 FT DOMAIN 52 784 GLN-RICH.
 FT DOMAIN 748 765 POLY-GLN.
 SQ SEQUENCE 865 AA; 96407 MW; A75499B2E9435C36 CRC64;

Query Match 35.6%; Score 42; DB 1; Length 865;
 Best Local Similarity 40.0%; Pred. No. 69;
 Matches 8; Conservative 4; Mismatches 8; Indels 0; Gaps 0;

OY 2 SPDSCEQQAARVQVHYKNE 21

```

Db      655 TPQHGYQHOOLOQQHKNEL 674
          :| : | : :| |||||
RESULT 14
TA29.TOBAC
ID TA29.TOBAC STANDARD; PRT; 321 AA.
AC P24804:
DT 01-MAR-1992 (Rel. 21, Created)
DT 01-MAR-1992 (Rel. 21, Last sequence update)
DT 01-MAY-1992 (Rel. 22, Last annotation update)
DE ANTHR-SPECIFIC PROTEIN TA-29.
GN GN TA-29.
OS Nicotiana tabacum (Common tobacco).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids I; Solanales; Solanaceae; Nicotiana.
NCBI_TaxId=4097;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90287727; PubMed=2356131;
RA Seurinck J., Truettner J., Goldberg R.B.;
RL "The nucleotide sequence of an anthr-specific gene.";
RT Nucleic Acids Res. 18:3403-3403(1990).
CC -|- TISSUE SPECIFICITY: ANTHR-SPECIFIC (TAPETAL CELLS).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to licensel@isb-sib.ch).
CC -----
CC EMBL; X52283; CAA36524.1; -
CC DR EMBL; A10061; CAA00875.1; -
CC FIR; S13550; S13550.
CC FT DOMAIN 52 58 POLY-GLY.
CC FT DOMAIN 67 77 POLY-GLY.
CC FT DOMAIN 206 215 POLY-GLY.
CC SEQUENCE 321 AA; 33389 MW; 9FE4AFB33F2070BD CRC64;
CC -----
Query Match 35.2%; Score 41.5; DB 1; Length 321;
Best Local Similarity 34.6%; Pred. No. 29;
Matches 9; Conservative 5; Mismatches 9; Indels 9; Gaps 1;
QY 3 PDSECEQARN-----VQHYHKN 19
    ||||| :| :| :| :|
DB 268 PDMCEDSCNELLHFVSPMQHKHEN 293
                                     :||| :| :|
RESULT 15
YNGC_ECOLI
ID YNGC_ECOLI STANDARD; PRT; 205 AA.
AC P76117;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE HYPOTHETICAL 23.3 KDA PROTEIN IN ANSP-RHSE INTERGENIC REGION.
GN YNGC OR B1454.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
NCBI_TaxId=562;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-K12 / MG1655;
RC MEDLINE=97426611; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;

```

RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence";
RL Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-CDMC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO FAMILY 65 OF GLYCOSYL HYDROLASES. STRONG,
CC TO M.LEPRAE ML0392.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z77165; CAB01022.1; .
DR EMBL; AE007156; AAK47846.1; .
DR TIGR; MT3509; .
DR TubercuList; RV3401; .
KW Hypothetical protein; Hydrolase; Glycosidase; Complete proteome.
SQ SEQUENCE 786 AA; 87315 MW; 73CB5E78DAD10AE6 CRC64;

Query Match 34.7%; Score 41; DB 1; Length 786;
Best Local Similarity 58.3%; Pred. No. 89;
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 7 EQARNYQVHYHK 18
DB 608 EQARNYDIYFER 619
||:||||:|

RESULT 18
YY01_MYCLE ID YY01_MYCLE STANDARD; PRT; 792 AA.
AC Q49736;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE HYPOPHYSICAL 88.1 KDA PROTEIN ML0392.
GN ML0392 OR B1620.F1.30.
OS Mycobacterium leprae.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1769;
RN [1]
RP SEQUENCE FROM N.A.
RA Smith D.R., Robinson K.;
RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=TN;
RX MEDLINE=21128732; PubMed=11234002;
RA Cole S.T., Eiglmeier K., Parkhill J., James K.D., Thomson N.R.,
RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,
RA Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,
RA Murphy L., Oliver K., Quail M.A., Rajandream M.-A., Rutherford K.M.,
RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,

RA Barrell B.G.;
 RT "Massive gene decay in the leprosy bacillus";
 RL Nature 409:1007-1011(2001).
 CC -1- SIMILARITY: BELONGS TO FAMILY 65 OF GLYCOSYL HYDROLASES. STRONG,
 CC TO M.TUBERCULOSIS RV3401.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; U00015; AAC43240.1; -;
 DR EMBL; AL583918; CAC29900.1; -;
 DR Leptoma; ML0392; -;
 KW Hypothetical protein; Hydrolase; Glycosidase; Complete proteome.
 SQ SEQUENCE 792 AA; 88087 MW; DFAFCBCCA4E20C8 CRC64;

 Query Match 34.7%; Score 41; DB 1; Length 792;
 Best Local Similarity 58.3%; Pred. No. 90;
 Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

 QY 7 EQQARNVQVHYK 18
 II:||||:|:
 DB 608 EQKARNVDYER 619

 RESULT 19
 ID CYGF_BOVIN STANDARD; PRT; 1103 AA.
 AC Q02740;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE RETINAL GUANYLYL CYCLASE 2 PRECURSOR (EC 4.6.1.2) (GUANYLATE CYCLASE
 DE 2F, RETINAL) (RETCG-2) (ROD OUTER SEGMENT MEMBRANE GUANYLATE CYCLASE
 DE 2) (ROS-GC2) (GUANYLATE CYCLASE F) (GC-F).
 GN GUCY2F OR GUC2F.
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovidae; Bovinae; Bos.
 OX NCBI_TaxID=9913;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-98238659; PubMed-9571173;
 RA Goraczniak R.M., Duda T., Sharma R.K.;
 RT "Calcium modulated signaling site in type 2 rod outer segment
 RT membrane guanylate cyclase (ROS-GC2).";
 RL Biochem. Biophys. Res. Commun. 245:447-453(1998).
 CC -1- FUNCTION: PROBABLY PLAYS A SPECIFIC FUNCTIONAL ROLE IN THE RODS
 CC AND/OR CONES OF PHOTORECEPTORS. IT MAY BE THE ENZYME INVOLVED IN
 CC THE RESYNTHESIS OF cGMP REQUIRED FOR RECOVERY OF THE DARK STATE
 CC AFTER PHOTOTRANSDUCTION (BY SIMILARITY).
 CC -1- CATALYTIC ACTIVITY: GTP -> 3',5'-CYCLIC GMP + PYROPHOSPHATE.
 CC -1- ENZYME REGULATION: ACTIVATED BY GCAP-1; INHIBITED BY CALCIUM.
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
 CC -1- PTM: THERE ARE 9 CONSERVED CYSTEINE RESIDUES IN SENSORY GUANYLATE
 CC CYCLASES, 6 IN THE EXTRACELLULAR DOMAIN, WHICH MAY BE INVOLVED IN
 CC INTRA- OR INTERCHAIN DISULFIDE BONDS.
 CC -1- SIMILARITY: BELONGS TO ADENYLYL CYCLASE CLASS-4/GUANYLYL CYCLASE
 CC FAMILY.
 CC -1- SIMILARITY: CONTAINS 1 PROTEIN KINASE-LIKE DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; U95958; IAB53864.1; -;
 DR HSSP; Q02846; IAWL.
 DR InterPro; IPR001828; ANF_receptor.
 DR InterPro; IPR000719; Euk_pkinase.
 DR InterPro; IPR001054; Guanylt_cyclase.
 DR Pfam; PF01094; ANF_receptor; 1.
 DR Pfam; PF00211; guanylate_cyc; 1.
 DR Pfam; PF00069; pkinase; 1.
 DR SMART; SM00044; CYCC; 1.
 DR PROSITE; PS00452; GUANYLATE_CYCLASES_1; 1.
 DR PROSITE; PS00125; GUANYLATE_CYCLASES_2; 1.
 DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
 DR KW Lyase; cGMP synthesis; Signal; Transmembrane; Multigene family;
 KW Vision.
 FT SIGNAL 1 46 BY SIMILARITY.
 FT CHAIN 47 1103 RETINAL GUANYLYL CYCLASE 2.
 FT DOMAIN 47 465 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 466 490 POTENTIAL.
 FT TRANSMEM 491 1103 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 532 812 PROTEIN KINASE-LIKE.
 FT DOMAIN 884 1014 GUANYLATE CYCLASE.
 FT DISULFID 104 132 BY SIMILARITY.
 FT DISULFID 452 452 INTERCHAIN (PROBABLE).
 FT DISULFID 460 460 INTERCHAIN (PROBABLE).
 SQ SEQUENCE 1103 AA; 124261 MW; EB731E1D8C642AA4 CRC64;

 Query Match 34.7%; Score 41; DB 1; Length 1103;
 Best Local Similarity 50.0%; Pred. No. 1.3e+02;
 Matches 10; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

 QY 2 SPDSCEQQAARNVQVHYK 21
 I:|||||:|:
 DB 374 SAASLVQHSRNVQVGFNQL 393

 RESULT 20
 ID CCAB_DISOM STANDARD; PRT; 2326 AA.
 AC P56698;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE PROBABLE VOLTAGE-DEPENDENT N-TYPE CALCIUM CHANNEL ALPHA-1B SUBUNIT
 DE (DOE-4).
 OS Discopoge ommata (Electric ray).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;
 OC Elasmobranchii; Squalia; Hyphosqualea; Pristiogalea; Batoidea;
 OC Torpediniformes; Narcinoidae; Narcinidae; Discopogidae.
 OX NCBI_TaxID=7785;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-Electric lobe;
 RX MEDLINE-93248175; PubMed-7683405;
 RA Horne W.A., Ellinor P.T., Inman I., Zhou M., Tsien R.W., Schwarz T.L.;
 RT "Molecular diversity of Ca2+ channel alpha 1 subunits from the marine
 RT ray Discopoge ommata";
 RL Proc. Natl. Acad. Sci. U.S.A. 90:3787-3791(1993).
 CC -1- FUNCTION: THE ISOFORM ALPHA-1B GIVES RISE TO N-TYPE CALCIUM
 CC CURRENTS. N-TYPE CALCIUM CHANNELS BELONG TO THE "HIGH-VOLTAGE
 CC ACTIVATED" (HVA) GROUP (BY SIMILARITY).
 CC -1- SUBUNIT: VOLTAGE-DEPENDENT CALCIUM CHANNELS ARE MULTISUBUNIT
 CC IN A 1:1:1 RATIO. THE CHANNEL ACTIVITY IS DIRECTED BY THE PORE-
 CC FORMING AND VOLTAGE-SENSITIVE ALPHA-1 SUBUNIT. IN MANY CASES, THIS
 CC SUBUNIT IS SUFFICIENT TO GENERATE VOLTAGE-SENSITIVE CALCIUM
 CC CHANNEL ACTIVITY. THE AUXILIARY SUBUNITS BETA AND ALPHA-2/DELTA
 CC LINKED BY A DISULFIDE BRIDGE REGULATE THE CHANNEL ACTIVITY (BY
 CC SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
 CC -1- ALTERNATIVE PRODUCTS: AT LEAST 2 ISOFORMS ARE PRODUCED BY


```
DR EMBL; U38804; AAC08104.1; -
KW Hypothetical protein; Chloroplast.
SQ SEQUENCE 450 AA; 52127 MW; 800F6FF8D5606C7C CRC64;

Query Match 34.3%; Score 40.5; DB 1; Length 450;
Best Local Similarity 53.8%; Pred. No. 60;
Matches 7; Conservative 5; Mismatches 0; Indels 1; Gaps 1;

QY 8 QOARNVQHYHKN 20
   |||:|:|:|:|:|
Db 292 QOAKNL-HFENQ 303

RESULT 22
MAD1_PETHY STANDARD; PRT; 231 AA.
AC Q07472;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DE FLORAL HOMEOTIC PROTEIN PMADS 1 (GREEN PETAL HOMEOTIC PROTEIN).
GN PMADS1 OR GP.
OS Petunia hybrida (petunia).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids I; Solanales; Solanaceae; Petunia.
OX NCBI_TaxID=4102;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Petal;
RX MEDLINE=94105323; PubMed=8278527;
RA Kush A., Brunelle A., Shevell D., Chua N.-H.;
RT "The cDNA sequence of two MADS box proteins in Petunia.";
RL Plant Physiol. 102:1051-1052(1993).
CC -!- FUNCTION: TRANSCRIPTION FACTOR INVOLVED IN THE GENETIC CONTROL OF
CC FLOWER DEVELOPMENT. NECESSARY FOR THE NORMAL DEVELOPMENT OF
CC PETALS. ABSENCE OF THE PMADS1 PROTEIN CAUSES TRANSFORMATION OF
CC PETALS INTO SEPALs.
CC -!- SUBCELLULAR LOCATION: NUCLEAR.
CC -!- TISSUE SPECIFICITY: PREDOMINANTLY EXPRESSED IN PETALS AND STAMENS,
CC LESS IN CARPELS AND SEPALs.
CC -!- SIMILARITY: BELONGS TO THE MADS DOMAIN FAMILY OF TRANSCRIPTION
CC FACTORS.
CC -!- SIMILARITY: CONTAINS A PROBABLE DIMERIZATION DOMAIN FOUND IN
CC SRF-TYPE TRANSCRIPTION FACTORS (K-BOX).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X69946; CAA49567.1; -
CC PIR; S31693; S31693.
CC HSSP; P11746; INNM.
CC TRANSFAC; T01777; -
CC -----
CC Mendel; 9804; PETHY; MADS.3.
CC InterPro; IPR002487; K-box.
CC InterPro; IPR002100; MADS-box.
CC Pfam; PF01486; K-box; 1.
CC Pfam; PF00319; SRF-TF; 1.
CC PRINTS; PR00404; MADSDOMAIN.
CC SMART; SM00432; MADS; 1.
CC PROSITE; PS00350; MADS_BOX_1; 1.
CC PROSITE; PS00666; MADS_BOX_2; 1.
CC Transcription regulation; DNA-binding; Activator; Nuclear protein;
CC Developmental protein.
FT DOMAIN 3 58 MADS.
FT DOMAIN 93 165 K-BOX.
SQ SEQUENCE 231 AA; 27039 MW; 76AE5E0B43488B6F CRC64;

Query Match 33.9%; Score 40; DB 1; Length 231;
Best Local Similarity 33.3%; Pred. No. 35;
Matches 7; Conservative 7; Mismatches 7; Indels 0; Gaps 0;

QY 1 NSPDSCEQOARNVQHYHKNEL 21
   |:::|::|::|:|
Db 145 NQIETFKKKVRNVEIHRNLL 165

RESULT 23
VASP_MOUSE STANDARD; PRT; 376 AA.
ID VASP_MOUSE
AC P70460; Q9R214;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DE VASODILATOR-STIMULATED PHOSPHOPROTEIN (VASP).
DE VASOILATOR-STIMULATED PHOSPHOPROTEIN (VASP).
GN VASP.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=129;
RX MEDLINE=96411679; PubMed=8812448;
RA Zimmer M., Fink T., Fischer L., Hauser W., Scherer K., Lichter P.,
RA Walter U.;
RT "Cloning of the VASP (vasodilator-stimulated phosphoprotein) genes in
RT human and mouse: structure, sequence, and chromosomal localization.";
RL Genomics 36:227-233(1996).
RN [2]
RP SEQUENCE FROM N.A., AND PHOSPHORYLATION.
RC STRAIN=C57BL/6J;
RX MEDLINE=99185054; PubMed=10085070;
RA Collins S.P., Uhler M.D.;
RT "Cyclic AMP- and cyclic GMP-dependent protein kinases differ in their
RT regulation of cyclic AMP response element-dependent gene
RT transcription.";
RL J. Biol. Chem. 274:8391-8404(1999).
CC -!- FUNCTION: ACTIN- AND PROFILIN-BINDING MICROFILAMENT-ASSOCIATED
CC PROTEIN. MAY ACT IN CONCERT WITH PROFILIN TO CONVEY SIGNAL
CC TRANSDUCTION TO ACTIN FILAMENT PRODUCTION (BY SIMILARITY).
CC -!- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: FOCAL ADHESIONS (BY SIMILARITY).
CC -!- PTM: MAJOR SUBSTRATE FOR CAMP-DEPENDENT (CAPK) AND CGMP-DEPENDENT
CC PROTEIN KINASE (CGPK) IN PLATELETS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X98475; CAA67108.1; -
CC EMBL; AF084548; AAD16045.1; -
CC MGD; MGI:109268; Vasp.
CC InterPro; IPR000697; RanBP1_VASP.
CC InterPro; IPR001960; WHI.
CC Pfam; PF00568; WHI; 1.
CC SMART; SM00461; WHI; 1.
CC Phosphorylation; Actin-binding.
FT DOMAIN 166 182 POLY-PRO.
FT DOMAIN 318 321 POLY-SER.
FT MOD_RES 153 153 PHOSPHORYLATION (BY CAPK AND CGPK) (BY
FT MOD_RES 235 235 PHOSPHORYLATION (BY CAPK AND CGPK) (BY
FT MOD_RES 274 274 PHOSPHORYLATION (BY CAPK AND CGPK) (BY
```



```

FT CONFLICT 209 209 SIMILARITY)
FT T -> A (IN REF. 2).
FT CONFLICT 288 288 MISSING (IN REF. 2).
SQ SEQUENCE 376 AA; 39824 MW; CC338D07519A0294 CRC64;

Query Match 33.9%; Score 40; DB 1; Length 376;
Best Local Similarity 41.2%; Pred. No. 59;
Matches 7; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

QY 2 SPDSCEQARNVQHYHK 18
DB 134 SPELEQKQKQPEHMER 150
II: ||| | : | :

RESULT 24
TIG_HAEN TIG_HAEN STANDARD; PRT; 432 AA.
AC P44837;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE TRIGGER FACTOR (TF).
GN TIG OR HI0713.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae.
OC Haemophilus.
OX NCBI_TaxID=727;
RN RP SEQUENCE FROM N.A.
RC STRAIN-RD / KW20 / ATCC 51907;
RX MEDLINE=9535630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus
RT influenzae Rd."
RL Science 269:496-512(1995).
CC -1- FUNCTION: INVOLVED IN PROTEIN EXPORT. ACTS AS A CHAPERONE BY
CC MAINTAINING THE NEWLY SYNTHESIZED PROTEIN IN AN OPEN CONFORMATION
CC (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE FKBP-TYPE PPIASE FAMILY. TIG SUBFAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U32754; AAC22370.1;
DR TIGR: HI0713;
DR InterPro: IPR001179; FKBP_PPIase.
DR Pfam: PF00254; FKBP; 1.
DR PROSITE: PS00453; FKBP_PPIASE.1; FALSE_NEG.
DR PROSITE: PS00454; FKBP_PPIASE.2; FALSE_NEG.
DR PROSITE: PS50059; FKBP_PPIASE.3; 1.
KW Cell division; Chaperone; Isomerase; Rotamase; Complete proteome.
FT DOMAIN 161 246 PPIASE, FKBP-TYPE.
FT SEQUENCE 432 AA; 48332 MW; C6DB71F502973096 CRC64;

Query Match 33.9%; Score 40; DB 1; Length 432;
Best Local Similarity 57.1%; Pred. No. 68;
Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 7 EQQARNVQHYHKN 20
DB 379 EQPAEVVEVYSKNE 392
II: | | : | | | | |

RESULT 25
SNX8_HUMAN SNX8_HUMAN STANDARD; PRT; 465 AA.
AC Q9V5X2;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE SORTING NEXIN 8.
GN SNX8.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN RP SEQUENCE FROM N.A.
RA Teasdale R.D., Gleeson P.A., Karlsson L.;
RT "Identification of eleven novel human sorting nexin molecules. A sub-
RT group of the sorting nexin family is associated with the early
RT endosomes."
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: MAY BE INVOLVED IN SEVERAL STAGES OF INTRACELLULAR
CC TRAFFICKING.
CC -1- SIMILARITY: BELONGS TO THE SORTING NEXIN FAMILY.
CC -1- SIMILARITY: CONTAINS 1 PHOX HOMOLOGY (PX) DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF121858; AAD27831.1;
DR InterPro: IPR001683; PX.
DR Pfam: PF00787; PX; 1.
DR SMART: SM00312; PX; 1.
DR Transport: Protein transport.
KW DOMAIN 72 185 PX.
FT SEQUENCE 465 AA; 52569 MW; 90C5EDB761C31E88 CRC64;

Query Match 33.9%; Score 40; DB 1; Length 465;
Best Local Similarity 38.9%; Pred. No. 74;
Matches 7; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

QY 4 DSCQEQARNVQHYHKNEL 21
DB 329 DLCERHERKGVHLKHQAL 346
II: | | : | | : |

RESULT 26
HDA2_CAEEL HDA2_CAEEL STANDARD; PRT; 507 AA.
AC Q09440;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PUTATIVE HISTONE DEACETYLASE C08B11.2.
GN C08B11.2.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Simms M.;

```

This SWISS-PROT entry is copyrighted. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announcement/ or send an email to licenses@isb-sib.ch).

EMBL; AF067164; AAD32448.1; -;
EMBL; AC021092; AAF24968.1; -;
HSSP; P08047; ISP2.
InterPro; IPR000822; znf-C2H2.
Pfam; PF00096; zf-C2H2; 20.
PRINTS; PR00048; ZINCINGER.
SMART; SM00355; ZnF_C2H2_18.
PROSITE; PS00028; ZINC_FINGER_C2H2_1; 17.
PROSITE; PS50157; ZINC_FINGER_C2H2_2; 18.
Transcription regulation; DNA-binding; Zinc-finger; Metal-binding;
Nuclear protein; Repeat.

	DOMAIN	1	91	KRNB.
FT	DOMAIN	92	618	ZINC FINGERS.
FT	ZN_FING	92	114	C2H2-TYPE.
FT	ZN_FING	120	142	C2H2-TYPE.
FT	ZN_FING	148	170	C2H2-TYPE.
FT	ZN_FING	176	198	C2H2-TYPE.
FT	ZN_FING	204	226	C2H2-TYPE.
FT	ZN_FING	232	254	C2H2-TYPE.
FT	ZN_FING	260	282	C2H2-TYPE.
FT	ZN_FING	288	310	C2H2-TYPE.
FT	ZN_FING	316	338	C2H2-TYPE.
FT	ZN_FING	344	366	C2H2-TYPE.
FT	ZN_FING	372	394	C2H2-TYPE.
FT	ZN_FING	400	422	C2H2-TYPE.
FT	ZN_FING	428	450	C2H2-TYPE.
FT	ZN_FING	456	478	C2H2-TYPE.
FT	ZN_FING	484	506	C2H2-TYPE.
FT	ZN_FING	512	534	C2H2-TYPE.
FT	ZN_FING	568	590	C2H2-TYPE.
FT	ZN_FING	596	618	C2H2-TYPE.
FT	CONFLICT	437	437	H -> Y (IN REF. 2).
FT	CONFLICT	556	556	E -> K (IN REF. 2).
FT	SEQUENCE	623 AA;	72374 MW;	DFFDF4741DAEBC7B CRC64;

Query Match 33.9%; Score 40; DB 1; Length 623;
Best Local Similarity 37.5%; Pred. No. le+02;
Matches 6; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

QY 6 CEOGARNVQHVKNEL 21
| : | | : | : | | | |
Db 47 CQTEARTICNSHKTEI 62

RESULT 28
ARP8_YEAST
ID ARP8_YEAST STANDARD; PRT; 881 AA.
AC Q12386;
DT DT
DI 01-NOV-1997 (Rel. 35, Created)
DE 01-NOV-1997 (Rel. 35, Last sequence update)
DE 15-JUL-1999 (Rel. 38, Last annotation update)
DE ACTIN-LIKE PROTEIN ARP8.
GN ARP8 OR YOR141C OR YOR3348C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
NCBI_TaxID=4932;
[1]
RN RP
RP SEQUENCE FROM N.A.
RX MEDLINE=97344368; PubMed=9200815;
RA Voss H., Benes V., Andrade M.A., Valencia A., Rechmann S., Teodoru C.,
RA Schwager C., Pages V., Sander C., Ansorge W.;
RT "DNA sequencing and analysis of 130 kb from yeast chromosome XV."
RL Yeast 13:655-672(1997).

```

RN  [2]
RP  GENE NAME.
RX  MEDLINE=97435478; PubMed=9290209;
RA  Poch O., Winsor B.;
RT  "Who's who among the Saccharomyces cerevisiae actin-related proteins?
RT  A classification and nomenclature proposal for a large family.";
RL  Yeast 13:1053-1058(1997).
CC  -1- SIMILARITY: BELONGS TO THE ACTIN FAMILY.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL; X94335; CAA64058.1; -
DR  EMBL; Z75043; CAA9341.1; -
DR  SGD; S0005667; ARP8.
DR  InterPro; IPR000279; Actin.
DR  SMART; SM00268; ACTIN; 1.
KW  Structural protein; Cytoskeleton.
FT  DOMAIN 22 27 POLY-ASP.
SQ  SEQUENCE 881 AA; 100208 MW; 8174851B6B077A19 CRC64;

Query Match 33.9%; Score 40; DB 1; Length 881;
Best Local Similarity 35.3%; Pred. No. 1.4e+02;
Matches 6; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

QY  4 DSCEQQARNVQHYHKE 20
   : 1:1:1:1:1:1:
DB  806 EELEEQHQNHQNGNE 822

RESULT 29
YKFO_YEAST
ID  YKFO_YEAST STANDARD; PRT; 922 AA.
AC  P35736;
DT  01-JUN-1994 (Rel. 29, Created)
DT  01-JUN-1994 (Rel. 29, Last sequence update)
DT  01-NOV-1997 (Rel. 35, Last annotation update)
DE  HYPOTHETICAL 103.1 KDA PROTEIN IN NUP120-CSE4 INTERGENIC REGION.
GN  YKLO50C OR YKL301 OR YKL263.
OS  Saccharomyces cerevisiae (Baker's yeast).
OC  Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC  Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX  NCBI_TaxID=4932;
RN  [1]
RP  SEQUENCE OF 1-452 FROM N.A.
RC  STRAIN=S288C;
RX  MEDLINE=94205268; PubMed=8154189;
RA  Purnelle B., Skala J., van Dyck L., Tettelin H., Goffeau A.;
RT  "The sequence of a 17.5 kb DNA fragment on the left arm of yeast
RT  chromosome XI identifies the protein kinase gene ELM1, the DNA
RT  primase gene PH12, a new gene encoding a putative histone and seven
RT  new open reading frames.";
RL  Yeast 9:1379-1384(1993).
RN  [2]
RP  SEQUENCE OF 410-922 FROM N.A.
RC  STRAIN=S288C;
RX  MEDLINE=94378723; PubMed=8091862;
RA  Rasmussen S.W.;
RT  "Sequence of a 28.6 kb region of yeast chromosome XI includes the
RT  FBAL and TOA2 genes, an open reading frame (ORF) similar to a
RT  translationally controlled tumour protein, one ORF containing motifs
RT  also found in plant storage proteins and 13 ORFs with weak or no
RT  homology to known proteins.";
RL  Yeast 10:S63-S68(1994).
CC  -1- SIMILARITY: TO YEAST YMR031C.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its

```

```

CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL; X75781; CAA53420.1; -
DR  EMBL; X71621; -; NOT_ANNOTATED_CDS.
DR  EMBL; Z28050; CAA81885.1; -
DR  PIR; S37871; S37871.
DR  SGD; S0001533; YKL050C.
KW  Hypothetical protein.
SQ  SEQUENCE 922 AA; 103143 MW; 362EE9A3F642DC8B CRC64;

Query Match 33.9%; Score 40; DB 1; Length 922;
Best Local Similarity 37.5%; Pred. No. 1.5e+02;
Matches 6; Conservative 6; Mismatches 2; Indels 2; Gaps 1;

QY  4 DSCEQQARNVQHYH 17
   1:1:1:1:1:1:
DB  547 DACERDLKNSIEHY 562

RESULT 30.
TCF8_HUMAN
ID  TCF8_HUMAN STANDARD; PRT; 1124 AA.
AC  P37275; Q13800; Q12924;
DT  01-OCT-1994 (Rel. 30, Created)
DT  15-JUL-1999 (Rel. 38, Last sequence update)
DT  20-AUG-2001 (Rel. 40, Last annotation update)
DE  TRANSCRIPTION FACTOR 8 (NIL-2-A ZINC FINGER PROTEIN) (NEGATIVE
DE  REGULATOR OF IL2).
GN  TCF8 OR AREB6.
OS  Homo sapiens (Human).
OC  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC  Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX  NCBI_TaxID=9606;
RN  [1]
RP  SEQUENCE FROM N.A.
RX  MEDLINE=94186507; PubMed=8138542;
RA  Watanabe Y., Kawakami K., Hirayama Y., Nagano K.;
RT  "Transcription factors positively and negatively regulating the Na,K-
RT  ATPase alpha 1 subunit gene.";
RL  J. Biochem. 114:849-855(1993).
RN  [2]
RP  SEQUENCE FROM N.A.
RX  Bachman N.J., Scarpulla R.C.;
RT  "A human zinc finger homeodomain protein homologous to the chicken
RT  delta-crystallin enhancer binding protein, delta EFL.";
RL  Submitted (JUL-1994) to the EMBL/GenBank/DBJ databases.
RN  [3]
RP  SEQUENCE OF 390-1124 FROM N.A.
RX  MEDLINE=92108424; PubMed=1840704;
RA  Williams T.M., Moolten D., Burlein J., Romano J., Bhaerman R.,
RA  Godillot A., Mellon M., Rauscher F.J. III, Kant J.A.;
RT  "Identification of a zinc finger protein that inhibits IL-2 gene
RT  expression.";
RL  Science 254:1791-1794(1991).
RN  [4]
RP  FUNCTION: INHIBITS INTERLEUKIN-2 (IL-2) GENE EXPRESSION. MAY BE
CC  RESPONSIBLE FOR TRANSCRIPTIONAL REPRESSION OF THE IL-2 GENE.
CC  ENHANCES OR REPRESSES THE PROMOTER ACTIVITY OF THE ATP1A1 GENE
CC  DEPENDING ON THE QUANTITY OF CDNA AND ON THE CELL TYPE.
CC  -1- SUBCELLULAR LOCATION: NUCLEAR.
CC  -1- TISSUE SPECIFICITY: EXPRESSED IN HEART AND SKELETAL MUSCLE, BUT
CC  NOT IN LIVER, SPLEEN, OR PANCREAS.
CC  -1- SIMILARITY: BELONGS TO DELTA-EFL/ZFP-1 FAMILY OF TWO-HANDED ZINC
CC  FINGER/HOMEODOMAIN PROTEINS.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its

```

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL; D15050; BAA03646.1; --
DR EMBL; U12170; AAA20602.1; --
DR EMBL; M81699; --; NOT_ANNOTATED_CDS.
DR HSSP; P08046; 1A1F.
CC MIM; 189909;
DR InterPro: IPR001356; Homeobox.
DR InterPro: IPR000822; Znf-C2H2.
DR Pfam; PF00096; zf-C2H2; 7.
DR PRINTS; PR00048; ZINC_FINGER.
DR SMART; SM00389; HOX; 1.
DR SMART; SM00355; ZNF_C2H2; 7.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 5.
DR PROSITE; PS00157; ZINC_FINGER_C2H2_2; 6.
KW Transcription regulation; DNA-binding; Nuclear protein; Zinc-finger;
KW Homeobox; Repressor; Activator; Metal-binding; Repeat.
FT ZN_FING 170 193
FT ZN_FING 200 222
FT ZN_FING 240 262
FT ZN_FING 268 292
FT ZN_FING 288 312
FT DNA_BIND 581 640
FT ZN_FING 904 926
FT ZN_FING 932 954
FT ZN_FING 960 981
FT DOMAIN 989 1124
FT CONFLICT 420 420
FT CONFLICT 609 609
FT CONFLICT 654 654
FT CONFLICT 672 672
FT CONFLICT 681 681
SQ SEQUENCE 1124 AA; 124073 MW; 0A2714CC37C848D1 CRC64;

Query Match 33.9%; Score 40; DB 1; Length 1124;
Best Local Similarity 43.8%; Pred. No. 1.9e+02;
Matches 7; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 4 DSCEDQARNVQHYHKN 19
DB 110 DECESDAENQNHDPN 125

RESULT 31
YNI7_YEAST STANDARD; PRT; 1178 AA.
AC P48231;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE HYPOTHETICAL 132.5 KDA PROTEIN IN TOP2-MKT1 INTERGENIC REGION.
GN YNL087W OR N2250.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=5288C / FY1679;
RX MEDLINE=96310628; PubMed=8740422;
RA Soler-Mira A., Saiz J.E., Ballesta J.P.G., Remacha M.;
RT "The sequence of a 17,933 bp segment of Saccharomyces cerevisiae
RT chromosome XIV contains the RHO2, TOP2, MKT1 and END3 genes and five
RT new open reading frames."
RL yeast 12:485-491(1996).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC -1- SIMILARITY: CONTAINS 3 C2 DOMAINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL; X89016; CAA61423.1; --
DR EMBL; Z1363; CAA95963.1; --
DR HSSP; P21707; IRSY.
DR SGD; S0005031; YNL087W.
DR InterPro: IPR000008; C2.
DR Pfam; PF00168; C2; 3.
DR SMART; SM00239; C2; 4.
DR PROSITE; PS00004; C2_DOMAIN_2; 3.
KW Hypothetical protein; Transmembrane; Repeat.
FT DOMAIN 380 463
FT DOMAIN 652 733
FT DOMAIN 985 1070
FT TRANSMEM 103 123
FT TRANSMEM 272 292
FT TRANSMEM 324 344
FT TRANSMEM 766 786
FT TRANSMEM 855 875
FT TRANSMEM 1134 1154
SQ SEQUENCE 1178 AA; 132509 MW; 6597DE191C739F41 CRC64;

Query Match 33.9%; Score 40; DB 1; Length 1178;
Best Local Similarity 66.7%; Pred. No. 2e+02;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 12 NVQHYHKN 20
DB 470 HANHYHKN 478

RESULT 32
ATC9_YEAST STANDARD; PRT; 1472 AA.
AC Q12697;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PROBABLE CATION-TRANSPORTING ATPASE 2 (EC 3.6.3.-).
GN YOR291W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RN SEQUENCE FROM N.A.
RA Cziepluch C., Jauniaux J.-C., Kordes E., Polrey R., Pujol A.,
RA Tobiasch E.;
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RN SEQUENCE OF 1-88 FROM N.A.
RA Cheret G., Sor F.;
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: ATP + H(2)O = ADP + PHOSPHATE.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- SIMILARITY: BELONGS TO THE CATION TRANSPORT ATPASES FAMILY (E1-E2
CC ATPASES). SUBFAMILY V.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL; Z75199; CAA99518.1; --
DR SGD; S0005817; YOR291W.

```

DR InterPro: IPR001757; E1-E2_ATPase.
DR InterPro: IPR001454; Hydrolase.
DR Pfam: PF00122; E1-E2_ATPase; 1.
DR Pfam: PF00702; Hydrolase; 1.
DR PROSITE: PS00154; ATPASE_E1_E2; 1.
KW Hypothetical protein; Hydrolase; Transmembrane; Phosphorylation;
KW Magnesium; ATP-binding.
FT DOMAIN 1 293
FT DOMAIN 294 315
FT DOMAIN 316 321
FT DOMAIN 322 344
FT DOMAIN 345 488
FT TRANSMEM 489 511
FT DOMAIN 512 514
FT TRANSMEM 515 533
FT DOMAIN 534 693
FT TRANSMEM 694 713
FT DOMAIN 714 726
FT TRANSMEM 727 748
FT DOMAIN 749 1244
FT TRANSMEM 1245 1264
FT DOMAIN 1265 1271
FT TRANSMEM 1272 1289
FT DOMAIN 1290 1307
FT TRANSMEM 1308 1331
FT DOMAIN 1332 1351
FT TRANSMEM 1352 1374
FT DOMAIN 1375 1387
FT TRANSMEM 1388 1407
FT DOMAIN 1408 1423
FT TRANSMEM 1424 1446
FT DOMAIN 1447 1472
FT MOD_RES 781 781
FT METAL 1187 1187
FT METAL 1191 1191
SQ SEQUENCE 1472 AA; 166749 MW; 6738AC22E561A4A9 CRC64;

Query Match 33.5%; Score 39.5; DB 1; Length 1472;
Best Local Similarity 40.9%; Pred. No. 3e+02;
Matches 9; Conservative 4; Mismatches 8; Indels 1; Gaps 1;

QY 1 NSPD-SCSEQQARNVQHYHKNEL 21
| | | | | | | | | | | | | | | | | | | | | |
DB 394 NYPNPSYEPNDINLSHHANEI 415

RESULT 33
YVAT_BACSU STANDARD; PRT; 148 AA.
AC P37504;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE HYPOTHETICAL 17.0 KDA PROTEIN IN CORF-TETB INTERGENIC REGION.
GN YVAT.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=96051385; PubMed=7584024;
RA Ogasawara N., Nakai S., Yoshikawa H.;
RT "Systematic sequencing of the 180 kilobase region of the Bacillus
RL subtilis chromosome containing the replication origin.";
RL DNA Res. 1:1-14(1994).
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; D26185; BAA05203.1; -
DR EMBL; Z99124; CAB16109.1; -
DR Subtilist; BG10031; YYAT.
DR InterPro; IPR000182; Acetyltransf_GCN5.
DR Pfam; PF00583; Acetyltransf; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 148 AA; 17003 MW; 72CDB9A50800A9F3 CRC64;

Query Match 33.1%; Score 39; DB 1; Length 148;
Best Local Similarity 29.4%; Pred. No. 32;
Matches 5; Conservative 8; Mismatches 4; Indels 0; Gaps 0;

QY 4 DSCCEQOARNVQHYHKNEL 20
| | | | | | | | | | | | | | | | | | | | | |
DB 39 DTLQEQCQHLYVHENQ 55

RESULT 34
YD54_AQUAE STANDARD; PRT; 150 AA.
ID YD54_AQUAE STANDARD; PRT; 150 AA.
AC O67367;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE HYPOTHETICAL PROTEIN AQ_1354.
GN AQ_1354.
OS Aquifex aeolicus.
OC Bacteria; Aquificales; Aquificaceae; Aquifex.
OX NCBI_TaxID=63363;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VF5;
RX MEDLINE=98196666; PubMed=9537320;
RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
RA Graham D.E., Overbeek R., Shead M.A., Keller M., Auja M., Huber R.,
RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
RT "The complete genome of the hyperthermophilic bacterium Aquifex
RL aeolicus.";
RL Nature 392:353-358(1998).
-----
-1- SIMILARITY: BELONGS TO THE UPF0054 FAMILY.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; AE000736; AAC07338.1; -
DR InterPro; IPR02036; UPF0054.
DR Pfam; PF02130; UPF0054; 1.
DR ProDom; PD005688; UPF0054; 1.
DR PROSITE; PS01306; UPF0054; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 150 AA; 17338 MW; 490C48DB6F952BCF CRC64;

Query Match 33.1%; Score 39; DB 1; Length 150;
Best Local Similarity 40.0%; Pred. No. 32;
Matches 8; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

QY 2 SPDCSEQOARNVQHYHKNEL 21
| | | | | | | | | | | | | | | | | | | | | |
DB 90 SQTAEQARELGHSEEV 109

RESULT 35

```


CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; X83276; CAA58263.1; -
 DR EMBL; Z74231; CAA98758.1; -
 DR SGD; S0002342; YDL183C.
 KW Hypothetical protein.
 SQ SEQUENCE 320 AA; 37047 MW; C287FD335D3D301 CRC64;

Query Match 33.1%; Score 39; DB 1; Length 320;
 Best Local Similarity 36.4%; Pred. No. 71;
 Matches 8; Conservative 4; Mismatches 6; Indels 4; Gaps 1;

OY 1 MSPDSCQQAQNVQH-----YHK 18
 :|||||:|:|:|
 Db 169 SSPDECLRMKKLYOGLIYHK 190

RESULT 38
 GRP2_MOUSE STANDARD; PRT; 322 AA.
 ID GRP2_MOUSE
 AC O89100;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE GRB2-RELATED ADAPTOR PROTEIN 2 (GADS PROTEIN) (GROWTH FACTOR RECEPTOR
 DE BINDING PROTEIN) (GRB2) (GRB-2-LIKE PROTEIN) (GRB2L) (HEMATOPOIETIC
 DE CELL-ASSOCIATED ADAPTOR PROTEIN GRPL) (GRB-2-RELATED MONOCYTIC ADAPTER
 DE PROTEIN) (MONOCYTIC ADAPTER) (MONA) (ADAPTER PROTEIN GRID).
 GS GRP2 OR MONA OR GADS OR GRB2L OR GRID.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 [1]
 RN SEQUENCE FROM N.A.
 RA Kedra D., Dumanski J.P.;
 RT "Cloning of the human and mouse growth factor receptor binding protein
 RT like genes.";
 RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
 RN [2]

RP SEQUENCE FROM N.A.
 RA Bourrette R.P., Arnaud S., Myles G.M., Rohrschneider L.R.,
 RA Blanchet J.P., Mouchiroud G.;
 RT "Mona, a novel hematopoietic-specific adaptor interacting with the
 RT macrophage-colony-stimulating factor receptor, is implicated in
 RT monocyte/macrophage development.";
 RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99087328; PubMed=9872323;
 RA Liu S.K., McGlade C.J.;
 RT "Gads is a novel SH2 and SH3 domain-containing adaptor protein that
 RT binds to tyrosine-phosphorylated Shc.";
 RL Oncogene 17:3073-3082(1998).
 RN [4]

RP SEQUENCE FROM N.A.
 RA Law C.-L., Ewings M.K., Chaudhary P.M., Solow S.A., Yun T.J.,
 RA Marshall A.J., Hood L., Clark E.A.;
 RT "Grpl, a Grb2-related adaptor protein, interacts with SLP-76 to
 RT regulate NF-AT activation.";
 RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RP SEQUENCE FROM N.A.
 RA Ellis J.H., Ashman C., Burden M., Kilpatrick K.E., Morse M.A.,
 RA Hamblin P.A.;
 RT "GRID, a novel Grb2-related adaptor protein which interacts with the

RT activated T cell co-stimulatory receptor CD28.";
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: INTERACTS WITH SLP-76 TO REGULATE NF-AT ACTIVATION.
 CC BINDS TO TYROSINE-PHOSPHORYLATED SHC.
 CC -!- SIMILARITY: CONTAINS 1 SH2 DOMAIN.
 CC -!- SIMILARITY: CONTAINS 2 SH3 DOMAINS.
 CC -!- SIMILARITY: BELONGS TO THE GRB2 / SEM-5 / DRK FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; AJ011735; CAA09756.1; -
 DR EMBL; AF055465; AAD08803.1; -
 DR EMBL; AF053405; AAC98669.1; -
 DR EMBL; AF129477; AAD41783.1; -
 DR EMBL; AF236118; AAF60318.1; -
 DR HSSP; P29354; IGR1.
 DR MGD; MGI:1333842; Mona.
 DR InterPro; IPR000980; SH2.
 DR InterPro; IPR001452; SH3.
 DR Pfam; PF00017; SH2; 1.
 DR Pfam; PF00018; SH3; 2.
 DR PRINTS; PR00452; SH3DOMAIN.
 DR SMART; SM00252; SH2; 1.
 DR SMART; SM00326; SH3; 2.
 DR PROSITE; PS50001; SH2; 1.
 DR PROSITE; PS50002; SH3; 2.
 KW SH2 domain; SH3 domain.
 FT DOMAIN 1 56 SH3 1.
 FT DOMAIN 58 149 SH2.
 FT DOMAIN 263 322 SH3 2.
 SQ SEQUENCE 322 AA; 36810 MW; 736311D0640CD3D0 CRC64;

Query Match 33.1%; Score 39; DB 1; Length 322;
 Best Local Similarity 50.0%; Pred. No. 72;
 Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;
 OY 8 QAARNVOHYHKN 19
 :|||:|:|:|
 Db 214 QQRYLQHFQD 225

RESULT 39
 WNT4_HUMAN STANDARD; PRT; 351 AA.
 ID WNT4_HUMAN
 AC P56705; Q9UJM2; Q9HJ8;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 20-AUG-2001 (Rel. 40, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE WNT-4 PROTEIN PRECURSOR.
 GN WNT4.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 [1]
 RN SEQUENCE FROM N.A.
 RP Testa T.T., Mossakowska D.E., Carter P.S., Hu E., Zhu Y.,
 RA Kelsall D.P., Murdock P.R., Herrity N.C., Lewis C.J., Cross D.A.,
 RA Culbert A.A., Reith A.D., Barnes M.R.;
 RT "Molecular cloning and characterization of six novel human WNT
 RT genes.";
 RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE OF 27-351 FROM N.A.
 RA Pearce A.;
 RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.

RP SEQUENCE OF 210-329 FROM N.A.
RX TISSUE-Breast;
RC MEDLINE-94221588; PubMed-8168088;
RA Huquet E.L., McMahon J.A., McMahon A.P., Bicknell R., Harris A.L.;
RT "Differential expression of human Wnt genes 2, 3, 4, and 7B in human
RT breast cell lines and normal and disease states of human breast
RT tissue.";
RL Cancer Res. 54:2615-2621(1994).
CC -!- FUNCTION: PROBABLE DEVELOPMENTAL PROTEIN. MAY BE A SIGNALING
CC MOLECULE WHICH AFFECT THE DEVELOPMENT OF DISCRETE REGIONS OF
CC TISSUES. IS LIKELY TO SIGNAL OVER ONLY FEW CELL DIAMETERS (BY
CC SIMILARITY). OVEREXPRESSION MAY BE ASSOCIATED WITH ABNORMAL
CC PROLIFERATION IN HUMAN BREAST TISSUE.
CC -!- SUBCELLULAR LOCATION: POSSIBLY SECRETED AND ASSOCIATES WITH THE
CC EXTRACELLULAR MATRIX.
CC -!- SIMILARITY: BELONGS TO THE WNT FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AY00398; AAC38658.1; -;
CC EMBL; AL031281; CAB52601.1; -;
CC MIM; 603490; -;
CC InterPro; IPR000970; Wnt1.
CC Pfam; PF00110; wnt; 1.
CC PRINTS; PR01349; WNTPROTEIN.
CC SMART; SM00097; WNT1.1.
CC PROSITE; PS00246; WNT1; 1.
CC Developmental protein; Glycoprotein; Signal.
CC SIGNAL 1 22 POTENTIAL.
CC CHAIN 23 351 WNT-4 PROTEIN.
CC CARBOHYD 88 88 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 297 297 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CONFLICT 106 106 I -> T (IN REF. 2).
CC CONFLICT 111 111 L -> F (IN REF. 2).
CC SEQUENCE 351 AA; 39029 MW; 25072318EDF1F93A CRC64;
CC
CC Query Match 33.1%; Score 39; DB 1; Length 351;
CC Best Local Similarity 63.6%; Pred. No. 79;
CC Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
CC
CC QY 2 SPDSCEQQAARN 12
CC III III I;
CC Db 276 SPDFCEQDMRS 286
CC
CC RESULT 40
CC WNT4_MOUSE
CC ID WNT4_MOUSE STANDARD; PRT; 351 AA.
CC AC P22724;
CC DT 01-AUG-1991 (Rel. 19, Created)
CC DT 01-AUG-1991 (Rel. 19, Last sequence update)
CC DT 15-JUL-1999 (Rel. 38, Last annotation update)
CC DE WNT-4 PROTEIN PRECURSOR.
CC GN WNT4 OR WNT-4.
CC OS Mus musculus (Mouse).
CC OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
CC NCBI_TaxID=10090;
CC [1]
CC SEQUENCE FROM N.A.
CC MEDLINE-91122634; PubMed-2279700;
CC Gavin B.J., McMahon J.A., McMahon A.P.;
CC "Expression of multiple novel Wnt-1/int-1-related genes during fetal
CC and adult mouse development.";
CC Genes Dev. 4:2319-2332(1990).
CC

CC -!- FUNCTION: MAY BE AN INTRACELLULAR SIGNALING MOLECULE INVOLVED IN
CC SEGMENTATION OF THE FOREBRAIN. IS LIKELY TO SIGNAL OVER ONLY FEW
CC CELL DIAMETERS (BY SIMILARITY). SEEMS TO BE INVOLVED IN KIDNEY
CC DEVELOPMENT.
CC -!- SUBCELLULAR LOCATION: POSSIBLY SECRETED AND ASSOCIATES WITH THE
CC EXTRACELLULAR MATRIX.
CC -!- TISSUE SPECIFICITY: IN ADULTS IN LUNG AND BRAIN.
CC -!- SIMILARITY: BELONGS TO THE WNT FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M89797; AAA40566.1; -;
CC PIR; C36470; C36470.
CC MGD; MGI:98957; Wnt4.
CC InterPro; IPR000970; Wnt1.
CC Pfam; PF00110; wnt; 1.
CC SMART; SM00097; WNT1; 1.
CC PROSITE; PS00246; WNT1; 1.
CC Developmental protein; Glycoprotein; Signal.
CC SIGNAL 1 22 POTENTIAL.
CC CHAIN 23 351 WNT-4 PROTEIN.
CC CARBOHYD 88 88 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 297 297 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC SEQUENCE 351 AA; 39049 MW; 7E1C5C739BE939D9 CRC64;
CC
CC Query Match 33.1%; Score 39; DB 1; Length 351;
CC Best Local Similarity 63.6%; Pred. No. 79;
CC Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
CC
CC QY 2 SPDSCEQQAARN 12
CC III III I;
CC Db 276 SPDFCEQDMRS 286
CC
CC RESULT 41
CC WNT4_RAT
CC ID WNT4_RAT STANDARD; PRT; 351 AA.
CC AC O9OX05;
CC DT 20-AUG-2001 (Rel. 40, Created)
CC DT 20-AUG-2001 (Rel. 40, Last sequence update)
CC DT 20-AUG-2001 (Rel. 40, Last annotation update)
CC DE WNT-4 PROTEIN PRECURSOR.
CC GN WNT4 OR WNT-4.
CC OS Rattus norvegicus (Rat).
CC OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
CC NCBI_TaxID=10116;
CC [1]
CC SEQUENCE FROM N.A.
CC STRAIN-SPRAGUE-DAWLEY; TISSUE-Corpus luteum;
CC Lacher M.D., Walther P.R., Lareu R., Dharmarajan A.M., Friis R.R.;
CC Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: MAY BE AN INTRACELLULAR SIGNALING MOLECULE INVOLVED IN
CC SEGMENTATION OF THE FOREBRAIN. IS LIKELY TO SIGNAL OVER ONLY FEW
CC CELL DIAMETERS (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: POSSIBLY SECRETED AND ASSOCIATES WITH THE
CC EXTRACELLULAR MATRIX.
CC -!- SIMILARITY: BELONGS TO THE WNT FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC


```

CC -----
DR EMBL: AF188608; AA15589.1; -
DR InterPro: IPR000970; Wnt1.
DR Pfam: PF00110; wnt; 1.
DR SMART: SM00097; WNT1; 1.
DR PROSITE: PS00246; WNT1; 1.
KW Developmental protein; Glycoprotein; Signal.
FT SIGNAL 1 22 POTENTIAL.
FT CHAIN 23 351 WNT-4 PROTEIN.
FT CARBOHYD 88 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 297 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 351 AA; 39043 MW; 5F8D80C3B4502BA1 CRC64;

Query Match 33.1%; Score 39; DB 1; Length 351;
Best Local Similarity 63.6%; Pred. No. 79;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 SPDSCEQOARN 12
Db ||| ||| |
276 SPDFCQDMRS 286

RESULT 42
ID WN8C_CHICK STANDARD; PRT; 357 AA.
AC P51030;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE WNT-8C PROTEIN PRECURSOR (CWNT-8).
GN WNT-8C.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=WHITE LEGHORN;
RX MEDLINE=94139558; PubMed=7916678;
RA Hume C.R., Dodd J.;
RT "Cwnt-8c": a novel wnt gene with a potential role in primitive streak
RT formation and hindbrain organization.";
RL Development 119:1147-1160(1993).
CC -!- FUNCTION: PROBABLE DEVELOPMENTAL PROTEIN. IS LIKELY TO SIGNAL OVER
CC ONLY FEW CELL DIAMETERS. MAY BE INVOLVED IN THE REGULATION OF AXIS
CC FORMATION AND IN THE RHOMBOMERE SPECIFICATION.
CC -!- SUBCELLULAR LOCATION: POSSIBLY SECRETED AND ASSOCIATES WITH THE
CC EXTRACELLULAR MATRIX.
CC -!- TISSUE SPECIFICITY: CELLS THAT FORM RHOMBOMERE 4. HENSEN'S NODE
CC AND THE NEURAL PLATE IMMEDIATELY ANTERIOR TO IT.
CC -!- DEVELOPMENTAL STAGE: EXPRESSED DURING EMBRYOGENESIS.
CC -!- SIMILARITY: BELONGS TO THE WNT FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).

-----
DR EMBL: U02097; AAA18933.1; -
DR InterPro: IPR000970; Wnt1.
DR Pfam: PF00110; wnt; 1.
DR SMART: SM00097; WNT1; 1.
DR PROSITE: PS00246; WNT1; 1.
KW Developmental protein; Glycoprotein; Signal.
FT SIGNAL 1 16 POTENTIAL.
FT CHAIN 17 357 WNT-8C PROTEIN.
FT CARBOHYD 104 104 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 263 263 N-LINKED (GLCNAC. . .) (POTENTIAL).

```

```

FT CARBOHYD 282 282 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 346 346 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 357 AA; 40042 MW; 58ADD7E835A5B8C4 CRC64;

Query Match 33.1%; Score 39; DB 1; Length 357;
Best Local Similarity 42.9%; Pred. No. 80;
Matches 9; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 1 NSPDSCEQOARNVQHYHKNEL 21
Db ||| ||| |
230 NSADSRGATAETFFHVHSTEL 250

RESULT 43
ID WN5B_ORYLA STANDARD; PRT; 371 AA.
AC O42122;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE WNT-5B PROTEIN PRECURSOR.
GN WNT5B.
OS Oryzias latipes (Medaka fish).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;
OC Belontiiformes; Adrianichthyidae; Oryziinae; Oryzias.
OX NCBI_TaxID=8090;
RN [1]
RP SEQUENCE FROM N.A.
RA Yokoi H., Nishimatsu A., Ozato K., Yoda K.;
RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: PROBABLE DEVELOPMENTAL PROTEIN. MAY BE A SIGNALING
CC MOLECULE WHICH AFFECT THE DEVELOPMENT OF DISCRETE REGIONS OF
CC TISSUES. IS LIKELY TO SIGNAL OVER ONLY FEW CELL DIAMETERS (BY
CC SIMILARITY).
CC -!- SUBCELLULAR LOCATION: POSSIBLY SECRETED AND ASSOCIATES WITH THE
CC EXTRACELLULAR MATRIX.
CC -!- SIMILARITY: BELONGS TO THE WNT FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).

-----
DR EMBL: AB006579; BAA22143.1; -
DR InterPro: IPR000970; Wnt1.
DR Pfam: PF00110; wnt; 1.
DR SMART: SM00097; WNT1; 1.
DR PROSITE: PS00246; WNT1; 1.
KW Developmental protein; Glycoprotein; Signal.
FT SIGNAL 1 28 POTENTIAL.
FT CHAIN 29 371 WNT-5B PROTEIN.
FT CARBOHYD 105 105 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 111 111 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 303 303 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 317 317 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 371 AA; 42280 MW; E9864E1FA342E82D CRC64;

Query Match 33.1%; Score 39; DB 1; Length 371;
Best Local Similarity 35.0%; Pred. No. 83;
Matches 7; Conservative 3; Mismatches 10; Indels 0; Gaps 0;

QY 1 NSPDSCEQOARNVQHYHKNEL 20
Db ||| ||| |
196 NYPRGSRHARTLMNLHNE 215

```



```

DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE INORGANIC PHOSPHATE TRANSPORTER PHO87.
GN PHO87 OR YCR037C OR YCR37C OR YCR524.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Schizosaccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / EV1679;
RX MEDLINE=91181345; PubMed=1964349;
RA Thierry A., Fairhead C., Dujon B.;
RT "The complete sequence of the 8.2 kb segment left of MAT on
RT chromosome III reveals five ORFs, including a gene for a yeast
RT ribokinase.";
RL Yeast 6:521-534 (1990).
RN [2]
RP CHARACTERIZATION.
RX MEDLINE=96171517; PubMed=8598055;
RA Bun-Ya M., Shikata K., Nakade S., Yompakdee C., Harashima S.,
RA Oshima Y.;
RT "Two new genes, PHO86 and PHO87, involved in inorganic phosphate
RT uptake in Saccharomyces cerevisiae.";
RL Curr. Genet. 29:344-351 (1996).
CC -!- FUNCTION: INVOLVED IN THE UPTAKE OF INORGANIC PHOSPHATE.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC -!- SIMILARITY: BELONGS TO THE NACD/P/PHO87 FAMILY OF TRANSPORTERS.
CC PHO87 SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X56909; CAA40229.1; -
DR EMBL; X59720; CAA42304.1; -
DR PIR; S12919; MMBY7C.
DR SGD; S0000633; PHO87.
KW Phosphate transport; Transmembrane.
FT TRANSMEM 462 482
FT TRANSMEM 494 514 POTENTIAL.
FT TRANSMEM 538 558 POTENTIAL.
FT TRANSMEM 584 604 POTENTIAL.
FT TRANSMEM 628 648 POTENTIAL.
FT TRANSMEM 668 688 POTENTIAL.
FT TRANSMEM 708 728 POTENTIAL.
FT TRANSMEM 736 756 POTENTIAL.
FT TRANSMEM 768 788 POTENTIAL.
FT TRANSMEM 803 823 POTENTIAL.
FT TRANSMEM 850 870 POTENTIAL.
FT TRANSMEM 899 919 POTENTIAL.
FT CARBOHYD 162 162 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 202 202 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 274 274 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 872 872 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 923 AA; 102549 MW; D168FC52FE26C5CB CRC64;

Query Match 33.1%; Score 39; DB 1; Length 923;
Best Local Similarity 42.9%; Pred. No. 2.2e+02;
Matches 6; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

OY 4 DSCQQQARNVQHYH 17
DB 375 DSCQKELKSLYLDH 388
||||:|:|:|

RESULT 46
YMLL_YEAST

DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE INORGANIC PHOSPHATE TRANSPORTER PHO87.
GN PHO87 OR YCR037C OR YCR37C OR YCR524.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Schizosaccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / EV1679;
RX MEDLINE=91181345; PubMed=1964349;
RA Thierry A., Fairhead C., Dujon B.;
RT "The complete sequence of the 8.2 kb segment left of MAT on
RT chromosome III reveals five ORFs, including a gene for a yeast
RT ribokinase.";
RL Yeast 6:521-534 (1990).
RN [2]
RP CHARACTERIZATION.
RX MEDLINE=96171517; PubMed=8598055;
RA Bun-Ya M., Shikata K., Nakade S., Yompakdee C., Harashima S.,
RA Oshima Y.;
RT "Two new genes, PHO86 and PHO87, involved in inorganic phosphate
RT uptake in Saccharomyces cerevisiae.";
RL Curr. Genet. 29:344-351 (1996).
CC -!- FUNCTION: INVOLVED IN THE UPTAKE OF INORGANIC PHOSPHATE.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC -!- SIMILARITY: BELONGS TO THE NACD/P/PHO87 FAMILY OF TRANSPORTERS.
CC PHO87 SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X56909; CAA40229.1; -
DR EMBL; X59720; CAA42304.1; -
DR PIR; S12919; MMBY7C.
DR SGD; S0000633; PHO87.
KW Phosphate transport; Transmembrane.
FT TRANSMEM 462 482
FT TRANSMEM 494 514 POTENTIAL.
FT TRANSMEM 538 558 POTENTIAL.
FT TRANSMEM 584 604 POTENTIAL.
FT TRANSMEM 628 648 POTENTIAL.
FT TRANSMEM 668 688 POTENTIAL.
FT TRANSMEM 708 728 POTENTIAL.
FT TRANSMEM 736 756 POTENTIAL.
FT TRANSMEM 768 788 POTENTIAL.
FT TRANSMEM 803 823 POTENTIAL.
FT TRANSMEM 850 870 POTENTIAL.
FT TRANSMEM 899 919 POTENTIAL.
FT CARBOHYD 162 162 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 202 202 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 274 274 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 872 872 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 923 AA; 102549 MW; D168FC52FE26C5CB CRC64;

Query Match 33.1%; Score 39; DB 1; Length 923;
Best Local Similarity 42.9%; Pred. No. 2.2e+02;
Matches 6; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

OY 4 DSCQQQARNVQHYH 17
DB 375 DSCQKELKSLYLDH 388
||||:|:|:|

RESULT 46
YMLL_YEAST

DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE INORGANIC PHOSPHATE TRANSPORTER PHO87.
GN PHO87 OR YCR037C OR YCR37C OR YCR524.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Schizosaccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RX Lye G., Churcher C.M., Barrell B.G., Rajandream M.A.;
RL Submitted (MAY-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 770-943 FROM N.A.
RA Pandit S., Sternglanz R.;
RL Submitted (DEC-1992) to the EMBL/GenBank/DBJ databases.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z49273; CAA89273.1; -
DR EMBL; L07650; AAA35122.1; -
DR PIR; S48514; S48514.
DR SGD; S0004731; YMR124W.
KW Hypothetical protein.
FT DOMAIN 59 66
FT DOMAIN 93 100 POLY-GLN.
FT DOMAIN 164 169 POLY-GLY.
FT DOMAIN 170 175 POLY-ASP.
FT DOMAIN 371 381 POLY-GLN.
FT DOMAIN 536 539 POLY-GLU.
FT DOMAIN 715 718 POLY-SER.
SQ SEQUENCE 943 AA; 105913 MW; 6631BEFBCD62E601 CRC64;

Query Match 33.1%; Score 39; DB 1; Length 943;
Best Local Similarity 41.2%; Pred. No. 2.2e+02;
Matches 7; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

OY 3 PDSCEQQARNVQHYHKN 19
DB 54 PSSVPQQQQQQQQYYRN 70
||||:|:|:|

RESULT 47
RPC1_SCHPO
ID RPC1_SCHPO STANDARD; PRT; 1405 AA.
AC O94666;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE DNA-DIRECTED RNA POLYMERASE III LARGEST SUBUNIT (EC 2.7.7.6).
GN RPC1 OR SPBC651.08C.
OS Schizosaccharomycetes pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RA Wood V., Rajandream M.A., Barrell B.G., Volckaert G.;
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION

```

CC OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS
 CC SUBSTRATES.
 CC -1- CATALYTIC ACTIVITY: N NUCLEOSIDE TRIPHOSPHATE - N PYROPHOSPHATE +
 CC RNA(N).
 CC -1- SUBUNIT: RNA POLYMERASE III CONSISTS OF ABOUT 15 DIFFERENT
 CC SUBUNITS. THIS SUBUNIT IS THE LARGEST COMPONENT OF RNA POLYMERASE
 CC III (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
 CC -1- MISCELLANEOUS: THREE DISTINCT ZINC-CONTAINING RNA POLYMERASES ARE
 CC FOUND IN EUKARYOTIC NUCLEI: POLYMERASE I FOR THE RIBOSOMAL RNA
 CC PRECURSOR, POLYMERASE II FOR THE MRNA PRECURSOR, AND POLYMERASE
 CC III FOR 5S AND TRNA GENES.
 CC -1- SIMILARITY: BELONGS TO THE RNA POLYMERASE BETA' CHAIN FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: AL035570; CAB37604.1; -
 CC InterPro: IPR000722; RNA_pol_A.
 CC DR InterPro: IPR002879; RNA_pol_A2.
 CC DR Pfam: PF00623; RNA_pol_A; 1.
 CC DR Pfam: PF01854; RNA_pol_A2; 1.
 CC DR Transferase: DNA-directed RNA polymerase; Transcription; zinc;
 CC zinc-finger; Nuclear protein.
 CC FT ZN_FING 66 79 C3H-TYPE (POTENTIAL).
 CC FT DOMAIN 1381 1385 POLY-GLU.
 CC SQ SEQUENCE 1405 AA; 157561 MW; 104B1AC4145A3B7B CRC64;
 CC
 CC Query Match 33.1%; Score 39; DB 1; Length 1405;
 CC Best Local Similarity 54.5%; Pred. No. 3.4e+02;
 CC Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
 CC
 CC QY 9 QARNVOHYHKN 19
 CC | |::|:|
 CC Db 808 QDRSLPHFHN 818
 CC
 CC RESULT 48
 CC RPOL_ASFB7
 CC ID RPOL_ASFB7 STANDARD; PRT; 1450 AA.
 CC AC P42486;
 CC DT 01-NOV-1995 (Rel. 32, Created)
 CC DT 01-NOV-1995 (Rel. 32, Last sequence update)
 CC DT 15-DEC-1998 (Rel. 37, Last annotation update)
 CC DE DNA-DIRECTED RNA POLYMERASE SUBUNIT 1 (EC 2.7.7.6).
 CC GN NP1450L.
 CC OS African swine fever virus (strain BA71V) (ASFV).
 CC OC Viruses; dsDNA viruses, no RNA stage; Asfarviridae;
 CC OC African swine fever-like viruses.
 CC OX NCBI_TaxID=10498;
 CC RN [1]
 CC RP SEQUENCE FROM N.A.
 CC RX MEDLINE=93281390; PubMed=8506138;
 CC RT Yanez R.J., Boursnell M., Nogal M.L., Yuste L., Vinuela E.;
 CC "African swine fever virus encodes two genes which share significant
 CC homology with the two largest subunits of DNA-dependent RNA
 CC polymerases.";
 CC RL Nucleic Acids Res. 21:2423-2427(1993).
 CC RN [2]
 CC RP SEQUENCE FROM N.A.
 CC RX Yanez R.J., Rodriguez J.M., Nogal M.L., Yuste L., Enriquez C.,
 CC Rodriguez J.F., Vinuela E.;
 CC "Analysis of the complete nucleotide sequence of African swine fever
 CC virus.";
 CC RL Virology 208:249-278(1995).
 CC -1- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION
 CC OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS

CC SUBSTRATES.
 CC -1- CATALYTIC ACTIVITY: N NUCLEOSIDE TRIPHOSPHATE - N PYROPHOSPHATE +
 CC RNA(N).
 CC -1- SIMILARITY: TO THE CORRESPONDING SUBUNITS OF PROKARYOTIC AND
 CC EUKARYOTIC RNA POLYMERASES.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: Z21489; CAA79697.1; -
 CC EMBL: U18486; AAA65328.1; -
 CC InterPro: IPR000722; RNA_pol_A.
 CC DR InterPro: IPR002879; RNA_pol_A2.
 CC DR Pfam: PF00623; RNA_pol_A; 1.
 CC DR Pfam: PF01854; RNA_pol_A2; 1.
 CC DR Transferase: DNA-directed RNA polymerase; Transcription.
 CC KW SEQUENCE 1450 AA; 163748 MW; 94D335C50B9A281B CRC64;
 CC SQ
 CC
 CC Query Match 33.1%; Score 39; DB 1; Length 1450;
 CC Best Local Similarity 57.1%; Pred. No. 3.5e+02;
 CC Matches 8; Conservative 1; Mismatches 5; Indels 0; Gaps 0;
 CC
 CC QY 8 QQARNVOHYHKNEL 21
 CC | |::|:|
 CC Db 349 QVAETQVHYNNRL 362
 CC
 CC RESULT 49
 CC COAC_SCHPO
 CC ID COAC_SCHPO STANDARD; PRT; 2280 AA.
 CC AC P78820; Q09447; Q09576; Q09667; Q09616; Q94557;
 CC DT 15-JUL-1998 (Rel. 36, Created)
 CC DT 20-AUG-2001 (Rel. 40, Last sequence update)
 CC DT 20-AUG-2001 (Rel. 40, Last annotation update)
 CC DE ACETYL-COA CARBOXYLASE (EC 6.4.1.2) (ACC) [INCLUDES: BIOTIN
 CC CARBOXYLASE (EC 6.3.4.14)].
 CC GN CUT6 OR SPAC56E4.04C.
 CC OS Schizosaccharomyces pombe (Fission yeast).
 CC OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 CC OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 CC OC Schizosaccharomycetes.
 CC OX NCBI_TaxID=4896;
 CC RN [1]
 CC RP SEQUENCE FROM N.A.
 CC RC STRAIN=HM123;
 CC RA Saito A., Kazuta Y., Toh H., Kondo H., Tanabe T.;
 CC Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
 CC RL [2]
 CC RP SEQUENCE FROM N.A.
 CC RC STRAIN=972;
 CC RA Connor R., Churcher C.M., Barrell B.G., Rajandream M.A., Wood V.;
 CC Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
 CC RN [3]
 CC RP SEQUENCE OF 14-161; 536-871; 998-1098 AND 1380-1547 FROM N.A.
 CC RX MEDLINE=96354912; PubMed=8769419;
 CC RA Saitoh S., Takahashi K., Nabeshima K., Yamashita Y., Nakaseko Y.,
 CC Hirata A., Yanagida M.;
 CC "Aberrant mitosis in fission yeast mutants defective in fatty acid
 CC synthetase and acetyl CoA carboxylase.";
 CC RL J. Cell Biol. 134:949-961(1996).
 CC -1- FUNCTION: THIS PROTEIN CARRIES THREE FUNCTIONS: BIOTIN CARBOXYL
 CC CARRIER PROTEIN, BIOTIN CARBOXYLASE, AND CARBOXYLTRANSFERASE.
 CC -1- CATALYTIC ACTIVITY: ATP + ACETYL-COA + HCO(3)(-) = ADP +
 CC ORTHOPHOSPHATE + MALONYL-COA.
 CC -1- CATALYTIC ACTIVITY: ATP + BIOTIN-CARBOXYL-CARRIER PROTEIN + CO(2)
 CC = ADP + ORTHOPHOSPHATE + CARBOXYBIOTIN-CARRIER PROTEIN.
 CC -1- COFACTOR: BIOTIN.

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 17, 2001, 07:49:38 ; Search time 12.81 Seconds
(without alignments)
124.876 Million cell updates/sec

Title: US-09-462-625-2_COPY_55_75
Perfect score: 118
Sequence: 1 NSPDSCEQQARNVQHYHKNEL 21

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 150 summaries

Database :

PIR_68:*
1: PIR1.*
2: PIR2.*
3: PIR3.*
4: PIR4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	48.5	41.1	523	2 T19584	hypothetical prote
2	46	39.0	222	2 S23731	MADS box protein T
3	45	38.1	139	2 T09878	albumin 2S storage
4	45	38.1	143	2 C69884	conserved hypotet
5	45	38.1	479	1 S41015	transcription fact
6	45	38.1	604	2 T36966	hypothetical prote
7	44	37.3	185	2 JC7369	prenylated Rab acc
8	44	37.3	227	2 S12378	MADS box protein d
9	44	37.3	358	2 T26281	hypothetical prote
10	44	37.3	471	2 S35635	DNA injection prot
11	44	37.3	640	2 S52047	deoxyribodipyrimid
12	44	37.3	734	2 T27055	hypothetical prote
13	44	37.3	778	2 H96649	protein F2401.4 [i
14	44	37.3	800	2 T01843	chloride channel p
15	44	37.3	945	2 S77052	cation-transportin
16	43	36.4	139	2 T09850	albumin 2S storage
17	43	36.4	170	2 T02707	probable kinetcho
18	43	36.4	327	2 T30072	hypothetical prote
19	43	36.4	347	2 A70180	spermidine/putresc
20	43	36.4	360	2 S34173	wnt-5c protein - A
21	43	36.4	440	2 B71858	adenylosuccinate 1
22	43	36.4	440	2 H64658	adenylosuccinate 1
23	43	36.4	452	2 T45448	probable serine pr
24	43	36.4	943	2 A42681	centromere protein
25	42.5	36.0	1002	2 C86026	probable oxidoredu
26	42.5	36.0	1166	1 E65151	hypothetical 126K
27	42	35.6	253	2 A65174	hypothetical 28.0
28	42	35.6	253	2 D86056	hypothetical prote
29	42	35.6	288	2 G75574	TDP-glucose-4,6-de

homeotic protein a
cell-cell signalin
hypothetical prote
adenylosuccinate 1
hypothetical prote
surface glycoprote
hypothetical prote
transcription acti
gastric mucin MUC5
hypothetical prote
hypothetical prote
anther-specific pr
hypothetical prote
hypothetical prote
hypothetical prote
hypothetical prote
extracellular prot
activin receptor S
DNA-binding protei
hypothetical coile
protein trihelix D
hypothetical prote
hypothetical prote
guanylate cyclase
N conserved hypoth
calcium channel pr
hypothetical prote
hypothetical prote
hypothetical prote
hypothetical prote
MADS box protein g
cytokine inducible
conserved hypotet
transport protein
hypothetical prote
conserved hypotet
methicillin resist
trigger factor tig
hypothetical prote
translation releas
methionine--trNA 1
hypothetical prote
hypothetical prote
hypothetical prote
hypothetical prote
probable membrane
hypothetical prote
cation-transportin
nuclear antigen EB
zinc finger protei
E-box-binding repr
probable membrane
hypothetical prote
hypothetical prote
BIR repeat contain
conserved hypotet
hypothetical prote
probable membrane
1-phosphatidylinos
probable phosphati
cyclic beta 1-2 gl
probable PE protei
conserved hypotet
conserved hypotet
hypothetical prote
hypothetical prote
probable membrane
MADS box protein h

103 39 33.1 228 2 T07066
104 39 33.1 240 2 H70248
105 39 33.1 276 2 T16590
106 39 33.1 271 2 E64567
107 39 33.1 284 2 T21923
108 39 33.1 304 2 D96652
109 39 33.1 318 2 S17197
110 39 33.1 320 2 S58734
111 39 33.1 325 2 D82662
112 39 33.1 338 2 E75301
113 39 33.1 351 2 C36470
114 39 33.1 357 2 I50690
115 39 33.1 410 2 D83573
116 39 33.1 424 2 T38271
117 39 33.1 435 2 S69035
118 39 33.1 480 2 A42383
119 39 33.1 507 2 T26809
120 39 33.1 603 2 F96797
121 39 33.1 854 1 ORHYLD
122 39 33.1 900 2 T14277
123 39 33.1 923 1 MMB7C
124 39 33.1 943 2 S54493
125 39 33.1 996 2 JE0237
126 39 33.1 1087 2 J01162
127 39 33.1 1405 2 T40607
128 39 33.1 1450 2 S78060
129 39 33.1 2279 2 T42531
130 39 33.1 2280 2 T38906
131 39 33.1 2562 2 T14266
132 38.5 32.6 66 2 G69174
133 38.5 32.6 196 2 T26943
134 38.5 32.6 207 2 A60969
135 38.5 32.6 221 2 S75111
136 38.5 32.6 440 2 I39847
137 38.5 32.6 463 2 C70931
138 38.5 32.6 601 2 T00119
139 38.5 32.6 645 2 T32820
140 38.5 32.6 1305 2 T31096
141 38.5 32.6 1505 2 T31418
142 38.5 32.6 2416 2 T13825
143 38 32.2 77 2 D71821
144 38 32.2 78 2 F82865
145 38 32.2 81 2 S24619
146 38 32.2 127 2 A48420
147 38 32.2 162 2 B27475
148 38 32.2 181 2 F83907
149 38 32.2 188 2 F71534
150 38 32.2 201 2 T50933

ALIGNMENTS

RESULT 1
T19684
hypothetical protein C33D9.5 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T19684
R:Lloyd, C.
submitted to the EMBL Data Library, December 1995
A:Reference number: Z19162
A:Accession: T19684
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-523 <WIL>
A:Cross-references: EMBL:Z68159; PIDN:CRA92288.1; GSPDB:GN00022; CESP:C33D9.5
A:Experimental source: clone C33D9
C:Genetics:
A:Gene: CESP:C33D9.5
A:Map position: 4
A:Introns: 165/1; 340/1; 459/1

Query Match 41.1%; Score 48.5; DB 2; Length 523;
Best Local Similarity 32.1%; Pred. No. 9.1;
Matches 9; Conservative 4; Mismatches 9; Gaps 1;
QY 3 PDSCEQQAARVQHY-----HKNEL 21
Db 195 PDECKNTRRIQHLTFIVSLQEAHRDEL 222
RESULT 2
S23731
MADS box protein TDR6 - tomato (fragment)
N:Alternate names: floral homeotic protein TM6
C:Species: Lycopersicon esculentum (tomato)
C>Date: 27-May-1994 #sequence_revision 26-May-1995 #text_change 26-Aug-1999
C:Accession: S23731; S38778
R:Pnueli, L.; Abu-Abdel, M.; Zamir, D.; Nacken, W.; Schwarz-Sommer, Z.; Lifschitz, E.
Plant J. 1, 255-266, 1991
A:Title: The MADS box gene family in tomato: temporal expression during floral develop
A:Reference number: S23728; MUID:93251098
A:Accession: S23731
A:Molecule type: mRNA
A:Residues: 1-222 <PNU>
A:Cross-references: EMBL:X60759
R:Pnueli, L.
submitted to the EMBL Data Library, July 1991
A:Reference number: S38778
A:Accession: S38778
A:Molecule type: mRNA
A:Residues: 1-159, 'T', 161-222 <PMF>
A:Cross-references: EMBL:X60759; NID:g19385; PID:g19386
C:Genetics:
A:Map position: 2
C:Superfamily: transcription factor squa; serum response factor DNA-binding domain ho
C:Keywords: DNA binding; nucleus; transcription regulation
F:1-54/Domain: serum response factor DNA-binding domain homology (fragment) <SRF>
Query Match 39.0%; Score 46; DB 2; Length 222;
Best Local Similarity 36.8%; Pred. No. 9;
Matches 7; Conservative 7; Mismatches 5; Indels 0; Gaps 0;
QY 1 NSPDSCEQQAARVQHYHKN 19
Db 142 NQPTCTCKKARNLEEQNGN 160
RESULT 3
T09878
albumin 2S storage protein precursor Mat5-A - upland cotton
C:Species: Gossypium hirsutum (upland cotton)
C>Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 29-Oct-1999
C:Accession: T09878
R:Galau, G.A.; Wang, H.Y.C.; Hughes, D.W.
submitted to the EMBL Data Library, January 1992
A:Description: Cotton Mat5-A (C164) gene and Mat5-D cDNAs encoding methionine-rich 2S
A:Reference number: Z16893
A:Accession: T09878
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-139 <GAL>
A:Cross-references: EMBL:M86213; NID:g167358; PID:g167359
C:Genetics:
A:Gene: Mat5-A
C:Keywords: storage protein
F:1-20/Domain: signal sequence #status predicted <SIG>
F:21-139/Product: albumin 2S storage protein Mat5-A #status predicted <MAT>
Query Match 38.1%; Score 45; DB 2; Length 139;
Best Local Similarity 45.8%; Pred. No. 7.9;
Matches 11; Conservative 1; Mismatches 6; Indels 6; Gaps 1;

C:Keywords: DNA binding
F:153-340/Domain: T-box homology <TBX>

Query Match 38.1%; Score 45; DB 1; Length 479;
Best Local Similarity 70.0%; Pred. No. 29;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 12 NVQHYHKNEL 21
| : ||||| :
Db 155 NLFYHKNEM 164

RESULT 6
T36966
hypothetical protein SCJ1.33 - Streptomyces coelicolor
C:Species: Streptomyces coelicolor
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C:Accession: T36966
R:Seeger, K.J.; Harris, D.; Thomson, N.R.; Parkhill, J.; Barrell, B.G.; Rajandream, M
submitted to the EMBL Data Library, September 1999
A:Reference number: 221607
A:Accession: T36966
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-604 <SEE>
A:Cross-references: EMBL:AL109962; PIDN:CAB53151.1; GSPDB:GN00070; SCOEDB:SCJ1.33
A:Experimental source: strain A3(2)
C:Genetics:
A:Gene: SCOEDB:SCJ1.33

Query Match 38.1%; Score 45; DB 2; Length 604;
Best Local Similarity 50.0%; Pred. No. 37;
Matches 8; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

Qy 3 PDSCQQARNVQHYHK 18
| : ||||| :
Db 484 PDVCAQGRGVTHRHE 499

RESULT 7
JC7369
prenylated Rab acceptor - mouse
C:Species: Mus musculus (house mouse)
C:Date: 17-Nov-2000 #sequence_revision 17-Nov-2000 #text_change 17-Nov-2000
C:Accession: JC7369
R:Liang, Z.; Li, G.
Biochem. Biophys. Res. Commun. 275, 509-516, 2000
A:Title: Mouse prenylated Rab acceptor is a novel Golgi membrane protein.
A:Reference number: JC7369
A:Contents: Brain
A:Accession: JC7369
A:Molecule type: mRNA
A:Residues: 1-185 <LIA>
A:Cross-references: GB:AF252856
C:Comment: This acceptor, a Rab-interacting Golgi integral membrane protein, function
ell.
C:Genetics:
A:Gene: pra
C:Keywords: brain; Golgi apparatus; membrane protein

Query Match 37.3%; Score 44; DB 2; Length 185;
Best Local Similarity 36.8%; Pred. No. 15;
Matches 7; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

Qy 1 NSPDSCQQARNVQHYHKN 19
| : ||||| :
Db 59 NVGELCQRLVRNVEYQSN 77

RESULT 8

Qy 4 DSCQQARNVQH-----YHKNEL 21
||||| | : ||
Db 34 DSCQQIRKQAHKHCQYMEEL 57

RESULT 4
C69884
conserved hypothetical protein ymca - Bacillus subtilis
C:Species: Bacillus subtilis
C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 15-Oct-1999
C:Accession: C69884
R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Bertex
C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Chd
A.; Ehrlich, S.D.; Emmeron, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
Nature 390, 249-256, 1997
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galleg
iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hulio, M.F.
Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinols
A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel
Y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon,
A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Sero
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpestra, P.; Tognoni, A.; Tosato, V.; Uchiyama,
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K
A:Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
A:Reference number: A69580; MUID:98044033
A:Accession: C69884
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-143 <KUN>
A:Cross-references: GB:Z99112; GB:AL009126; MID:g2633902; PIDN:CAB13575.1; PID:ell185293;
A:Experimental source: strain 168
C:Genetics:
A:Gene: ymca

Query Match 38.1%; Score 45; DB 2; Length 143;
Best Local Similarity 40.0%; Pred. No. 8.2;
Matches 8; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

Qy 1 NSPDSCQQARNVQHYHKN 20
| : ||||| :
Db 47 NQIRALQKQAVNLKHYKHE 66

RESULT 5
S41015
transcription factor tbx8 - Caenorhabditis elegans
N:Alternate names: hypothetical protein T07C4.2
C:Species: Caenorhabditis elegans
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C:Accession: S41015; A58530
R:Berk, M.

submitted to the EMBL Data Library, January 1994
A:Reference number: S41014
A:Accession: S41015
A:Molecule type: DNA
A:Residues: 1-479 <BER>
A:Cross-references: EMBL:Z29443
R:Aguilnik, S.I.; Bollag, R.J.; Silver, L.M.
Genomics 25, 214-219, 1995
A:Title: Conservation of the T-box gene family from Mus musculus to Caenorhabditis eleg
A:Reference number: A56530; MUID:95293375
A:Accession: A56530
A:Status: nucleic acid sequence not shown; not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 153-340 <AGU>
A:Cross-references: GB:Z29443
C:Genetics:
A:Introns: 77/3; 101/3; 134/3; 236/2; 261/3; 312/3; 450/3
C:Superfamily: Caenorhabditis elegans transcription factor tbx8; T-box homology

S35635
DNA injection protein - phase P22
N;Alternate names: gp20 protein
C;Species: phage P22
C;Date: 22-Nov-1993 #sequence_revision 12-May-1995 #text_change 08-Oct-1999
C;Accession: S35635; D43330
R;Adhikari, P.; Berget, P.B.
Nucleic Acids Res. 21, 1499, 1993
A;Title: Sequence of a DNA injection gene from Salmonella typhimurium phase P22.
A;Reference number: S35635; MUID:93219140
A;Accession: S35635
A;Molecule type: DNA
A;Residues: 1-471 <ADH>
A;Cross-references: EMBL:L07556; NID:g215274; PIDN:AAA62407.1; PID:g215276
R;Conlin, C.A.; Vimr, E.R.; Miller, C.G.
J. Bacteriol. 174, 5869-5880, 1992
A;Title: Oligopeptidase A is required for normal phase P22 development.
A;Reference number: A43330; MUID:92394890
A;Contents: Salmonella typhimurium
A;Accession: D43330
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-92 <CON>
A;Cross-references: GB:M93985; NID:g215283; PIDN:AAA72116.1; PID:g215286
A;Note: sequence extracted from NCBI backbone (NCBIN:113001, NCBIP:113005)

Query Match 37.3%; Score 44; DB 2; Length 471;
Best Local Similarity 36.8%; Pred. No. 41;
Matches 7; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

QY 1 NSPDSCEQQARNVQHYNK 19
:||||:| | | | |
DB 411 SSPEAMQOSIREIOEYTNN 429

RESULT 11
S52047
deoxyribodipyrimidine photo-lyase (EC 4.1.99.3) - fruit fly (Drosophila melanogaster)
N;Alternate names: DNA photolyase; Photoreactivating enzyme
C;Species: Drosophila melanogaster
C;Date: 15-Jul-1995 #sequence_revision 01-Sep-1995 #text_change 20-Jun-2000
C;Accession: S52047
R;Yasui, A.; Eker, A.P.M.; Yasuhira, S.; Vajima, H.; Kobayashi, T.; Takao, M.; EMBO J. 13, 6143-6151, 1994
A;Title: A new class of DNA photolyases present in various organisms including A;Reference number: S52046; MUID:95112825
A;Accession: S52047
A;Status: nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-640 <AS>
A;Cross-references: EMBL:D26021; NID:gl401029; PIDN:BAA05042.1; PID:gl401030
C;Genetics:
A;Gene: FlyBase:phr
A;Cross-references: FlyBase:FBgn0003082
C;Superfamily: deoxyribodipyrimidine photo-lyase
C;Keywords: carbon-carbon lyase

Query Match 37.3%; Score 44; DB 2; Length 640;
Best Local Similarity 33.3%; Pred. No. 56;
Matches 7; Conservative 6; Mismatches 8; Indels 0; Gaps 0;

QY 1 NSPDSCEQQARNVQHYNK 21
| : ||||| : :
DB 15 NEINNLRQVRNLQHVQRKD 35

RESULT 12
T27055
hypothetical protein Y49E10.23 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999

[illegible]

F:21-139/Product: albumin 2S storage protein #status predicted <MAT>

Query Match 36.4%; Score 43; DB 2; Length 139;
 Best Local Similarity 66.7%; Pred. No. 16;
 Matches 8; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 4 DSCFQOQARNVQH 15
 ||||| | |
 DB 34 DSCFQOQIRKQAH 45

RESULT 17
 T02707
 Probable kinetochore (Skp1p-like) protein At2g03190 [imported] - Arabidopsis thaliana
 N; Alternate names: SKP1-like protein T18E12.14
 C: Species: Arabidopsis thaliana (mouse-ear cress)
 C: Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 16-Feb-2001
 C: Accession: T02707; D84445
 R: Rounsley, S.D.; Lin, X.; Kaul, S.; Shea, T.P.; Fujii, C.Y.; Mason, T.M.; Shen, M.; Ron
 submitted to the EMBL Data Library, September 1998
 A: Description: Arabidopsis thaliana chromosome II BAC T18E12 genomic sequence.
 A: Reference number: Z14702
 A: Accession: T02707
 A: Status: translated from GB/EMBL/DBJ
 A: Molecule type: DNA
 A: Residues: 1-170 <ROU>
 A: Cross-references: EMBL:AC005313; NID:g3548797; PID:g3548811
 A: Experimental source: cultivar Columbia
 M.; Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
 M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.;
 euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
 Nature 402, 761-768, 1999
 A: Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
 A: Reference number: A84420; MUID:20083487
 A: Accession: D84445
 A: Status: preliminary
 A: Molecule type: DNA
 A: Residues: 1-170 <STO>
 A: Cross-references: GB:AE002093; NID:g3548811; PIDN:AAC34483.1; GSPDB:GN00139
 C: Genetics:
 A: Gene: At2g03190; T18E12.14
 A: Map position: 2
 C: Superfamily: human S-phase kinase-associated protein 1A

Query Match 36.4%; Score 43; DB 2; Length 170;
 Best Local Similarity 47.6%; Pred. No. 20;
 Matches 10; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

QY 1 NSPDSCEQOQARNVQHYHKNEL 21
 :||| | | |
 DB 69 DSDSDSTATSENVNEEARNEL 89

RESULT 18
 T30072
 hypothetical protein F57F4.2 - Caenorhabditis elegans
 C: Species: Caenorhabditis elegans
 C: Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
 C: Accession: T30072
 R: Miller, N.; Bradshaw, H.
 submitted to the EMBL Data Library, September 1996
 A: Description: The sequence of C. elegans cosmid F57F4.
 A: Reference number: Z20730
 A: Accession: T30072
 A: Status: preliminary; translated from GB/EMBL/DBJ
 A: Molecule type: DNA
 A: Residues: 1-377 <WIL>
 A: Cross-references: EMBL:U70856; PIDN:AB09165.1; GSPDB:GN00023; CESP:F57F4.2
 A: Experimental source: strain Bristol N2; clone F57F4
 C: Genetics:
 A: Gene: CESP:F57F4.2

A: Map position: 5
 A: Introns: 231/2

Query Match 36.4%; Score 43; DB 2; Length 327;
 Best Local Similarity 58.3%; Pred. No. 40;
 Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 8 QOARNVQHYHKN 19
 :| | | | | | | | | |
 DB 201 RKAASVAHYHKN 212

RESULT 19
 A70180
 spermidine/putrescine ABC transporter, ATP-binding protein (potA) homolog - Lyme dise
 C: Species: Borrelia burgdorferi (Lyme disease spirochete)
 C: Date: 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change 02-Feb-2001
 C: Accession: A70180
 R: Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; Wh
 son, D.; Peterson, J.; Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vu
 ; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.
 Nature 390, 580-585, 1997
 A: Authors: Smith, H.O.; Venter, J.C.
 A: Title: Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi.
 A: Reference number: A70100; MUID:98065943
 A: Accession: A70180
 A: Status: preliminary; nucleic acid sequence not shown; translation not shown
 A: Molecule type: DNA
 A: Residues: 1-347 <KLE>
 A: Cross-references: GB:AE001165; GB:AE000783; NID:g2688561; PIDN:AAB91525.1; PID:g268
 A: Experimental source: strain B31
 C: Superfamily: unassigned ATP-binding cassette proteins; ATP-binding cassette homolog
 C: Keywords: ATP; nucleotide binding; P-loop
 F: 23-214/Domain: ATP-binding cassette homolog <ABC>
 F: 40-47/Region: nucleotide-binding motif A (P-loop)

Query Match 36.4%; Score 43; DB 2; Length 347;
 Best Local Similarity 37.5%; Pred. No. 42;
 Matches 6; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

QY 4 DSCFQOQARNVQHYHKN 19
 | | | | | | | | | |
 DB 2 DNCILEIKNLSHYYDN 17

RESULT 20
 S34173
 wnt-5c protein - African clawed frog
 C: Species: Xenopus laevis (African clawed frog)
 C: Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 16-Jul-1999
 C: Accession: S34173; S45242
 R: Koster, J.G.; Kuiken, G.A.; Stegeman, B.; Peterson, J.; Eizema, K.; Stabel, L.; Dek
 submitted to the EMBL Data Library, June 1993
 A: Description: Differential Xwnt-5c expression during early development of xenopus la
 A: Reference number: S34173
 A: Accession: S34173
 A: Molecule type: mRNA
 A: Residues: 1-360 <KOS>
 A: Cross-references: EMBL:X73510; NID:g313267; PIDN:CAA51916.1; PID:g313268
 R: Kuiken, G.A.; Bertens, P.J.A.; Peterson-Maduro, J.; Veenstra, G.J.C.; Koster, J.G.;
 Nucleic Acids Res. 22, 1675-1680, 1994
 A: Title: The promoter of the Xwnt-5c gene contains octamer and AP-2 motifs functional
 A: Reference number: S45242; MUID:94261437
 A: Accession: S45242
 A: Molecule type: DNA
 A: Residues: 1-28 <KUI>
 C: Superfamily: int-1 transforming protein

Query Match 36.4%; Score 43; DB 2; Length 360;
 Best Local Similarity 40.0%; Pred. No. 44;

```
Query Match          36.4%; Score 43; DB 2; Length 440;
Best Local Similarity 41.2%; Pred. No. 54;
Matches 7; Conservative 5; Mismatches 4; Indels 0; Caps 0;

Qy      5 SCEQQARNVQHVKNEL 21
       ||| : | : : | : : |
Db     232 SCEKIAVNIRHLQRSEV 248

RESULT 23
T45448
```

RESULT 25
C86026
probable oxidoreductase subunit yhjL [imported] - Escherichia coli (strain O157:H7)
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 31-Mar-2001
C:Accession: C86026
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamouisis, K.; Apoda
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A95480; MUID:21074935; PMID:11206551
A:Accession: C86026
A>Status: preliminary
A:Molecule type: DNA

A;Residues: 1-1002 <STO>
A;Cross-references: GB:AE005174; NID:gl2518248; PIDN:AAG58671.1; GSPDB:GN00145; UWGP:Z49
A;Experimental source: strain O157:H7, substrain EDL933
C;Genetics:
A;Gene: yhjL

Query Match 36.0%; Score 42.5; DB 2; Length 1002;
Best Local Similarity 55.6%; Pred. No. 1.5e+02;
Matches 10; Conservative 2; Mismatches 3; Indels 3; Gaps 1;

QY 1 NSPDSCE---QQARNVOH 15
| : | | | | :
Db 210 NNPDRAERLFQQARNVDN 227
| : | | | | :

RESULT 26
E65151
hypothetical protein (dcta-gppF intergenic region) - Escherichia coli (strain K-12)
N;Alternate names: hypothetical protein flf65
C;Species: Escherichia coli
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 21-Jul-2000
C;Accession: E65151; S47751
R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Cohen, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A;Title: The complete genome sequence of Escherichia coli K-12.
A;Reference number: A64720; MUID:97426617
A;Accession: E65151
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-1166 <BLAT>
A;Cross-references: GB:AE000430; GB:U00096; NID:g2367238; PIDN:AAC76555.1; PID:g2367239;
A;Experimental source: strain K-12, substrain MG1655
R;Plunkett, G.
submitted to the EMBL Data Library, March 1994
A;Reference number: S47666
A;Accession: S47751
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-602, 'V', 605-1166 <PLU>
A;Cross-references: EMBL:U00039; NID:g466582; PID:g912469
C;Genetics:
A;Gene: yhjL
A;Start codon: GTG
C;Superfamily: hypothetical protein flf65; tetratricopeptide repeat homology
F:252-285/Domain: tetratricopeptide repeat homology <Tn1>
F:286-319/Domain: tetratricopeptide repeat homology <Tt2>

Query Match 36.0%; Score 42.5; DB 1; Length 1166;
Best Local Similarity 55.6%; Pred. No. 1.8e+02;
Matches 10; Conservative 2; Mismatches 3; Indels 3; Gaps 1;

QY 1 NSPDSCE---QQARNVOH 15
| : | | | | :
Db 374 NNPDRAERLFQQARNVDN 391
| : | | | | :

RESULT 27
A65174
hypothetical 28.0 kD protein in tnaB-bglB intergenic region - Escherichia coli (strain K-12)
C;Species: Escherichia coli
C;Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 08-Oct-1999
C;Accession: A65174
R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Cohen, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A;Title: The complete genome sequence of Escherichia coli K-12.
A;Reference number: A64720; MUID:97426617
A;Accession: A65174
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA

A;Residues: 1-253 <BLAT>
A;Cross-references: GB:AE000448; GB:U00096; NID:gl1790142; PIDN:AAC76735.1; PID:gl17901
A;Experimental source: strain K-12, substrain MG1655
C;Genetics:
A;Gene: yieE

Query Match 35.6%; Score 42; DB 2; Length 253;
Best Local Similarity 35.3%; Pred. No. 44;
Matches 6; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

QY 3 PDSCEQQAARNVQHYNK 19
| : | | | | :
Db 25 PSQCHOEARNIIPPHRS 41
| : | | | | :

RESULT 28
D86056
hypothetical protein yieE [imported] - Escherichia coli (strain O157:H7)
C;Species: Escherichia coli
C;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 31-Mar-2001
C;Accession: D86056
R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimallanta, E.; Potamouisis, K.; Apoda, Nature 409, 529-533, 2001
A;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A;Reference number: A85480; MUID:21074935; PMID:11206551
A;Accession: D86056
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-253 <STO>
A;Cross-references: GB:AE005174; NID:gl2518557; PIDN:AAG58912.1; GSPDB:GN00145; UWGP:
A;Experimental source: strain O157:H7, substrain EDL933
C;Genetics:
A;Gene: yieE

Query Match 35.6%; Score 42; DB 2; Length 253;
Best Local Similarity 35.3%; Pred. No. 44;
Matches 6; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

QY 3 PDSCEQQAARNVQHYNK 19
| : | | | | :
Db 25 PSQCHOEARNIIPPHRS 41
| : | | | | :

RESULT 29
G75574
TDP-glucose-4,6-dehydratase-related protein - Deinococcus radiodurans (strain R1)
C;Species: Deinococcus radiodurans
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 31-Mar-2000
C;Accession: G75574
R;White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.; Smith, H.O.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; S.; Shen, M.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A;Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A;Reference number: A75250; MUID:20036896
A;Accession: G75574
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-288 <WHI>
A;Cross-references: GB:AE001863; GB:AE001825; NID:g6460670; PIDN:AAF12411.1; PID:g646
A;Experimental source: strain R1
C;Genetics:
A;Gene: DRA0222
A;Map position: 2

Query Match 35.6%; Score 42; DB 2; Length 288;
Best Local Similarity 47.1%; Pred. No. 50;
Matches 8; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

A;Residues: 1-1002 <STO>
A;Cross-references: GB:AE005174; NID:gl2518248; PIDN:AAG58671.1; GSPDB:GN00145; UWGP:Z49
A;Experimental source: strain O157:H7, substrain EDL933
C;Genetics:
A;Gene: yhjL

Query Match 36.0%; Score 42.5; DB 2; Length 1002;
Best Local Similarity 55.6%; Pred. No. 1.5e+02;
Matches 10; Conservative 2; Mismatches 3; Indels 3; Gaps 1;

QY 1 NSPDSCE---QQARNVOH 15
|:||| | :||||| :
Db 210 NNPDRAERLFQQARNVDN 227

RESULT 26
E65151
hypothetical protein (dcta-gppF intergenic region) - Escherichia coli (strain K-12)
N;Alternate names: hypothetical protein flf65
C;Species: Escherichia coli
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 21-Jul-2000
C;Accession: E65151; S47751
R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Cohen, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A;Title: The complete genome sequence of Escherichia coli K-12.
A;Reference number: A64720; MUID:97426617
A;Accession: E65151
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-1166 <BLAT>
A;Cross-references: GB:AE000430; GB:U00096; NID:g2367238; PIDN:AAC76555.1; PID:g2367239;
A;Experimental source: strain K-12, substrain MG1655
R;Plunkett, G.
submitted to the EMBL Data Library, March 1994
A;Reference number: S47666
A;Accession: S47751
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-602,'V',605-1166 <PLU>
A;Cross-references: EMBL:U00039; NID:g466582; PID:g912469
C;Genetics:
A;Gene: yhjL
A;Start codon: GTG
C;Superfamily: hypothetical protein flf65; tetratricopeptide repeat homology
F:252-285/Domain: tetratricopeptide repeat homology <Tn1>
F:286-319/Domain: tetratricopeptide repeat homology <Tn2>

Query Match 36.0%; Score 42.5; DB 1; Length 1166;
Best Local Similarity 55.6%; Pred. No. 1.8e+02;
Matches 10; Conservative 2; Mismatches 3; Indels 3; Gaps 1;

QY 1 NSPDSCE---QQARNVOH 15
|:||| | :||||| :
Db 374 NNPDRAERLFQQARNVDN 391

RESULT 27
A65174
hypothetical 28.0 kD protein in tnaB-bglB intergenic region - Escherichia coli (strain K-12)
C;Species: Escherichia coli
C;Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 08-Oct-1999
C;Accession: A65174
R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Cohen, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A;Title: The complete genome sequence of Escherichia coli K-12.
A;Reference number: A64720; MUID:97426617
A;Accession: A65174
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA

A;Residues: 1-253 <BLAT>
A;Cross-references: GB:AE000448; GB:U00096; NID:gl1790142; PIDN:AAC76735.1; PID:gl17901
A;Experimental source: strain K-12, substrain MG1655
C;Genetics:
A;Gene: yieE

Query Match 35.6%; Score 42; DB 2; Length 253;
Best Local Similarity 35.3%; Pred. No. 44;
Matches 6; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

QY 3 PDSCEQQAARNVQHYNK 19
| | | | | : : :
Db 25 PSQCHOEARNIIPPHRS 41

RESULT 28
D86056
hypothetical protein yieE [imported] - Escherichia coli (strain O157:H7)
C;Species: Escherichia coli
C;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 31-Mar-2001
C;Accession: D86056
R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimallanta, E.; Potamouisis, K.; Apoda, Nature 409, 529-533, 2001
A;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A;Reference number: A85480; MUID:21074935; PMID:11206551
A;Accession: D86056
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-253 <STO>
A;Cross-references: GB:AE005174; NID:gl2518557; PIDN:AAG58912.1; GSPDB:GN00145; UWGP:
A;Experimental source: strain O157:H7, substrain EDL933
C;Genetics:
A;Gene: yieE

Query Match 35.6%; Score 42; DB 2; Length 253;
Best Local Similarity 35.3%; Pred. No. 44;
Matches 6; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

QY 3 PDSCEQQAARNVQHYNK 19
| | | | | : : :
Db 25 PSQCHOEARNIIPPHRS 41

RESULT 29
G75574
TDP-glucose-4,6-dehydratase-related protein - Deinococcus radiodurans (strain R1)
C;Species: Deinococcus radiodurans
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 31-Mar-2000
C;Accession: G75574
R;White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.; Smith, H.O.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; S.; Shen, M.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A;Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A;Reference number: A75250; MUID:20036896
A;Accession: G75574
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-288 <WHI>
A;Cross-references: GB:AE001863; GB:AE001825; NID:g6460670; PIDN:AAF12411.1; PID:g646
A;Experimental source: strain R1
C;Genetics:
A;Gene: DRA0222
A;Map position: 2

Query Match 35.6%; Score 42; DB 2; Length 288;
Best Local Similarity 47.1%; Pred. No. 50;
Matches 8; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 4 DSC EQ AR NV QH YK NE 20
| | : | | : | |
Db 232 DD A ER F AA EE V EH I H GNE 248

RESULT 30
S58850
homeotic protein abd-A - Junonia coenia
N:Alternate names: abdominal-A homeodomain protein
C:Species: Junonia coenia
C:Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 17-Nov-2000
C:Accession: S58850; S58851
R:Warren, R.W.; Nagy, L.; Selegue, J.; Gates, J.; Carroll, S.
Nature 372, 459-461, 1994
A:Title: Evolution of homeotic gene regulation and function in flies and butterflies.
A:Reference number: S58850; MUID:95075456
A:Accession: S58850
A>Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: mRNA
A:Residues: 1-295 <WAR>
A:Cross-references: EMBL:L41931; NID:G97276; PIDN:AAA68460.1; PID:G97277
A>Note: the nucleotide sequence was submitted to the EMBL Data Library, May 1995
A:Accession: S58851
A>Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: mRNA
A:Residues: 214-271 <WA2>
A:Cross-references: EMBL:L42135; NID:G833750; PIDN:AAA68461.1; PID:G833751
A>Note: the nucleotide sequence was submitted to the EMBL Data Library, May 1995
C:Genetics:
A:Gene: abd-a
C:Superfamily: unassigned homeobox proteins; homeobox homology
C:Keywords: DNA binding; homeobox; nucleus; transcription regulation
F:215-271/Domain: homeobox homology<HOX>

Query Match 35.6%; Score 42; DB 2; Length 295;
Best Local Similarity 42.9%; Pred. No. 51;
Matches 6; Conservative 4; Mismatches 4; Indels

```
QY 4 DSCEQQARNVQHYH 17
      : | : | | | : :
Db 38 EGCDQQLRPAQHYY 51
```

RESULT 31
 B56349
 cell-cell signaling molecule Awnt-5B precursor - axolotl
 C:Species: Ambystoma mexicanum (axolotl)
 C:Date: 21-Jul-1995 #sequence_revision 21-Jul-1995 #text_change 16-Jul-1999
 C:Accession: B56349; S25000
 R:Busse, U.; Segulin, C.
 Mech. Dev. 40, 63-72, 1993
 A:Title: Isolation of cDNAs for two closely related members of the axolotl Wnt family, A
 A:Reference number: A56349; MUID:33183769
 A:Accession: B56349
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-357 <BUS>
 A:Cross-references: EMBL:214048; NID:g62428; PIDN:CAA78416.1; PID:g62429
 A:Experimental source: embryo
 A:Note: sequence extracted from NCBI backbone (NCBIP:126896)
 C:Superfamily: int-1 transforming protein

Query Match 35.6%; Score 42; DB 2; Length 357;
Best Local Similarity 40.0%; Pred. No. 62;
Matches 8; Conservative 3; Mismatches 9; Indels

Qy 1 NSPDSCEQARNVQHYHKNE 20
| | | : : |
Db 182 NYPKGSCEQARTLMNLONNE 201

RESULT
T34467
32

hypothetical protein zk770.3 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:date: 23-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 02-Sep-2000
 C:Accession: T34467
 R:Maggi, L.; Gattung, S.; Bartko, L.
 submitted to the EMBL Data Library, April 1997
 A:Description: The sequence of C. elegans cosmid zk770.
 A:Reference number: Z21530

A:Accession: J34407
A:Status: preliminary; translated from GB/EMBL/DBJ/J
A:Molecule type: DNA
A:Molecule type: DNA
A:Residues: 1-408 <MAG>
A:Cross-references: EMBL:U97404; PIDN:AAB93310.1; G
A:Experimental source: strain Bristol N2; clone ZK7
C:Genetics:
A:Gene: CESP:ZK770.3
A:Map position: 1
A:Map position: 1
A:Introns: 43/2; 141/1; 202/3; 326/3
C:Superfamily: Passover protein

Query Match	35.6%	Score 42;	DB 2;	Length 408;
Best Local Similarity	37.5%;	Pred. No. 72;		
Matches	6;	Conservative	5;	Mismatches
			5;	Indels
			0;	Gaps
			0;	

```
QY      6 CEQARNVQHYHKNEL 21
      || : || : || :
Db     256 CEYEVRLDNIHKHSV 271
```

RESULT 33

B81418
adenylosuccinate lyase (EC 4.3.2.2) Cj0223 [imported] - Campylobacter jejuni (strain
C;Species: Campylobacter jejuni
C;Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-May-2000
C;Accession: B81418
R:Parkhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Basham, D.; Chillingworth,
C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVleet, A.; Whitehead, S.; Barrell,
N.A.; 403, 665-668, 2000 .
A;Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals
A;Reference number: AB1250; MUID:20150912

A/Accession: B61418
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-442 <PAR>
A/Cross-references: GB:AL111168; GB:AL139074; NID:g6967505; PIDN:CA872516.1; PID:g6967506
A/Experimental source: serotype O2, strain NCTC 11168
C/Genetics:

Query Match	35.6%	Score 42;	DB 2;	Length 442;
Best Local Similarity	41.2%	Pred. No. 78;		
Matches	7;	Conservative	5;	Mismatches
			5;	Indels
				Gaps

```
QY      5 SCEQQARNVQHYHKNEL 21
      ||||| :|: :|:
Db     232 SCEQIAVAIRHEORTEV 248
```

RESULT 34

R23021 34
 T25934
 hypothetical protein W02C12.3 - *Caenorhabditis elegans*
 C:Species: *Caenorhabditis elegans*
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
 C:Accession: T25934
 R:Murray, J.; Wohlmann, P.
 submitted to the EMBL Data Library, December 1996
 A:Description: The sequence of *C. elegans* cosmid W02C12.

A:Reference number: Z20112
A:Accession: T25934
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-498 <MUR>
A:Cross-references: EMBL:U80815; PIDN:AAB37997.1; GSPDB:GN000022; CESP:W02C12.3
A:Experimental source: strain Bristol N2; clone W02C12
C:Genetics:
A:Gene: CESP:W02C12.3
A:Map position: 4
A:Introns: 96/3; 128/3; 153/3; 196/3; 245/1; 270/2; 366/3; 417/1; 442/3

Query Match 35.6%; Score 42; DB 2; Length 498;
Best Local Similarity 50.0%; Pred. No. 89;
Matches 8; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 2 SPDSCEQQARNVQHYH 17
||| | | | | |
Db 479 SPDPINTQOSNGHYH 494

RESULT 35
QOXPGB
Surface glycoprotein polyprotein - Pichinde virus
C:Species: Pichinde virus
C:Date: 28-Aug-1985 #sequence_revision 28-Aug-1985 #text_change 23-Jul-1999
C:Accession: A04149
R:Auperlin, D.D.; Romanowski, V.; Galinski, M.; Bishop, D.H.L.
J. Virol. 52, 897-904, 1984
A:Title: Sequencing studies of Pichinde arenavirus S RNA indicate a novel coding strategy
A:Reference number: A93005; MUID:85033957
A:Accession: A04149
A:Molecule type: genomic RNA
A:Residues: 1-503 <AUP>
A:Cross-references: GB:K02734; NID:G332643; PIDN:AAA46824.1; PID:G332644
C:Comment: The genome consists of two species of RNA, designated S (small) RNA and L (large) RNA. The S RNA is the coding strand. The L RNA is the noncoding strand. The S RNA is the coding strand. The L RNA is the noncoding strand.
C:Superfamily: arenavirus surface glycoprotein
C:Keywords: glycoprotein
F:67,74,89,100,111,116,121,132,181,241,379,387,404,409,499/Binding site: carbohydrate (A

Query Match 35.6%; Score 42; DB 1; Length 503;
Best Local Similarity 44.4%; Pred. No. 90;
Matches 8; Conservative 2; Mismatches 6; Indels 2; Gaps 1;

QY 1 NSPDSCEQQARNVQHYH 18
| | | | | | | | | |
Db 81 NLPQSCSK--NNTHHYK 96

RESULT 36
D84601
hypothetical protein At2g21450 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:Accession: D84601
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umavam, L.; Tallon, L.;
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487
A:Accession: D84601
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-816 <STO>
A:Cross-references: GB:AE002093; NID:g4567276; PIDN:AAD23689.1; GSPDB:GN00139
C:Genetics:
A:Gene: At2g21450
A:Map position: 2

Query Match 35.6%; Score 42; DB 2; Length 816;
Best Local Similarity 54.5%; Pred. No. 1.5e+02;
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 8 QOARNVQHYH 18
: | | | | |
Db 521 ENTRNLHYH 531

RESULT 37
T46651
transcription activator protein acu-15 [imported] - Neurospora crassa
C:Species: Neurospora crassa
C:Date: 18-Feb-2000 #sequence_revision 18-Feb-2000 #text_change 02-Sep-2000
C:Accession: T46651
R:Bibbins, M.; Connerton, I.F.
submitted to the EMBL Data Library, March 1997
A:Description: Characterization of the regulator of acetate induction from N.Crassa.
A:Reference number: Z23116
A:Accession: T46651
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-865 <BIB>
A:Cross-references: EMBL:Y11565; PIDN:CAA72324.1
A:Experimental source: strain 74A (Str)
C:Genetics:
A:Gene: acu-15
A:Introns: 140/2; 150/1; 349/2
C:Superfamily: GAL4 zinc binuclear cluster homology
F:19-56/Domain: GAL4 zinc binuclear cluster homology <GL4>

Query Match 35.6%; Score 42; DB 2; Length 865;
Best Local Similarity 40.0%; Pred. No. 1.6e+02;
Matches 8; Conservative 4; Mismatches 8; Indels 0; Gaps 0;

QY 2 SPDSCEQQARNVQHYHKNEL 21
: | | | | | | | | | |
Db 655 TPQHQQYQHQQLOOQHKNEL 674

RESULT 38
JE0095
gastric mucin MUC5AC precursor - human
C:Species: Homo sapiens (man)
C:Date: 22-May-1998 #sequence_revision 29-May-1998 #text_change 07-May-1999
C:Accession: JE0095
R:van de Bovenkamp, J.H.B.; Hau, C.M.; Strous, G.J.A.M.; Buehler, H.A.; Dekker, J.; E
Biochem. Biophys. Res. Commun. 245, 853-859, 1998
A:Title: Molecular cloning of human gastric mucin MUC5AC reveals conserved cysteine-r
F:1-19/Domain: signal sequence #status predicted <SIG>
F:273-300/Domain: leucine zipper #status predicted <L2P>

Query Match 35.6%; Score 42; DB 2; Length 1373;
Best Local Similarity 43.8%; Pred. No. 2.6e+02;
Matches 7; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 3 PDSCEQQARNVQHYH 18
| | | | | | | | | |
Db 333 PDFCPQKCPNNQYHE 348

RESULT 39
T22945
hypothetical protein T01D3.1 - Caenorhabditis elegans


```
Query Match      35.2%; Score 41.5; DB 2; Length 463;
Best Local Similarity 55.6%; Pred. No. 98;
Matches 10; Conservative 1; Mismatches 4; Indels 3; Gaps 1;

QY 2 SPDSCEQQARNVQHYHKN 19
:||||| | | | |
Db 391 NPDSCLEQQ---QDYIKN 405

RESULT 44
T32980
hypothetical protein T21D12.7 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C:Accession: T32980
R:Du, Z.; Maggi, L.
submitted to the EMBL Data Library, February 1998
A:Description: The sequence of C. elegans cosmid K02D7.
A:Reference number: Z21259
A:Accession: T32980
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-692 <DU>
A:Cross-references: EMBL:AF045645; PIDN:AA02606.1; GSPDB:GN00022; CESP:T21D12.7
A:Experimental source: strain Bristol N2; clone K02D7
C:Genetics:
A:Gene: CESP:T21D12.7
A:Map position: 4
A:Introns: 29/1; 89/1; 127/1; 186/1; 233/1; 291/1; 379/3; 586/3; 611/3; 643/3

Query Match      35.2%; Score 41.5; DB 2; Length 692;
Best Local Similarity 45.0%; Pred. No. 1.5e+02;
Matches 9; Conservative 4; Mismatches 2; Indels 5; Gaps 1;

QY 1 NSPDSC-----QQARNVOH 15
:| ||| | | | | | |
Db 362 SSRDSCPEQYMQQANLEH 381

RESULT 45
C82468
hypothetical protein VCA0364 VCA0292 [imported] - Vibrio cholerae (strain N16961 serogroup O1)
C:Species: Vibrio cholerae
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
C:Accession: C82468; D82476
R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.;
chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, B.
I., R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A:Reference number: A82035; MUID:20406833
A:Accession: C82468
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-88 <HEI>
A:Cross-references: GB:AE003853; NID:99657758; PIDN:AAF96271.1; GSPDB:GN001
A:Experimental source: serogroup O1; strain N16961; biotype El Tor
A:Accession: D82476
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-88 <HE2>
A:Cross-references: GB:AE004369; GB:AE003853; NID:99657685; PIDN:AAF96200.1; GSPDB:GN001
A:Experimental source: serogroup O1; strain N16961; biotype El Tor
C:Genetics:
A:Gene: VCA0364; VCA0292
A:Map position: 2

Query Match      34.7%; Score 41; DB 2; Length 88;
Best Local Similarity 50.0%; Pred. No. 21;
Matches 7; Conservative 4; Mismatches 3; Indels 3; Gaps 0;

QY 3 PDSCEQQARNVQHYHKN 16
||| | | | | | | |
Db 23 PDSEQQKRNKKHF 36

RESULT 46
C83946
hypothetical protein BH2371 [imported] - Bacillus halodurans (strain C-125)
C:Species: Bacillus halodurans
C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 31-Dec-2000
C:Accession: C83946
R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; H.
Nucleic Acids Res. 28, 4317-4331, 2000
A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans a
A:Reference number: A83650; MUID:20263314
A:Accession: C83946
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-146 <STO>
A:Cross-references: GB:AP001515; GB:BA000004; NID:gl0174886; PIDN:BA06090.1; GSPDB:G
A:Experimental source: strain C-125
C:Genetics:
A:Gene: BH2371

Query Match      34.7%; Score 41; DB 2; Length 146;
Best Local Similarity 50.0%; Pred. No. 35;
Matches 7; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 7 EQQARNVQHYHKN 20
::: | | | | | |
Db 54 OKEAVNLQHYGKKE 67

RESULT 47
A64898
hypothetical protein bl454 - Escherichia coli
C:Species: Escherichia coli
C:Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 17-Mar-2000
C:Accession: A64898
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;
.A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617
A:Accession: A64898
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-205 <BLAT>
A:Cross-references: GB:AE000242; GB:U00096; NID:gl787720; PIDN:AAC74536.1; PID:gl7877
A:Experimental source: strain K-12, substrain MG1655
C:Superfamily: glutathione transferase

Query Match      34.7%; Score 41; DB 2; Length 205;
Best Local Similarity 44.4%; Pred. No. 50;
Matches 8; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY 2 SPDSCEQQARNVQHYHKN 19
:| | | | | | | |
Db 118 APDAPEQLKKNVIEYRKS 135

RESULT 48
RGSXD
extracellular proteinase response regulator degU - Bacillus subtilis
N:Alternate names: extracellular proteinase regulatory protein iep
C:Species: Bacillus subtilis
C:Date: 30-Jun-1991 #sequence_revision 30-Jun-1991 #text_change 16-Jun-2000
C:Accession: C30191; B30190; A31097; C69614
R:Henner, D.J.; Yang, M.; Ferrari, E.
J. Bacteriol. 170, 5102-5109, 1988
A:Title: Localization of Bacillus subtilis sacU(Hy) mutations to two linked genes wit
```

A:Reference number: A30191; MUID:89033891
A:Accession: C30191
A:Molecule type: DNA
A:Residues: 1-229 <HEN>
A:Cross-references: GB:M23558; NID:g143497; PIDN:AAA22733.1; PID:g143499
R:Kunst, F.; Debarbouille, M.; Meadek, T.; Young, M.; Maueel, C.; Karamata, D.; Klier, A.
J. Bacteriol. 170, 5093-5101, 1988
A:Title: Deduced polypeptides encoded by the Bacillus subtilis sacU locus share homology
A:Reference number: A30190; MUID:89033890
A:Accession: B30190
A:Molecule type: DNA
A:Residues: 1-229 <KUN>
A:Cross-references: GB:M23649; NID:g143500; PIDN:AAA22735.1; PID:g143502
R:Tanaka, T.; Kawata, M.
J. Bacteriol. 170, 3593-3600, 1988
A:Title: Cloning and characterization of Bacillus subtilis iep, which has positive and n
A:Reference number: A31097; MUID:88298669
A:Accession: A31097
A:Molecule type: DNA
A:Residues: 1-229 <TAN>
A:Cross-references: GB:M21658; NID:g143087; PIDN:AAA22545.1; PID:g143089
A:Experimental source: strain 168
R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berton
C.; Bron, S.; Brouillet, S.; Bruchli, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Chd
A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
Nature 390, 249-256, 1997
A:Authors: Foulger, D.; Fritz, C.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gallen
teck, J.; Harwood, C.R.; Henaut, A.; Hibert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.
Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois
A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel
Y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon
A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seron
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K
A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.
A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
A:Reference number: A69580; MUID:98044033
A:Accession: C69614
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-229 <RU2>
A:Cross-references: GB:Z99122; GB:AL009126; NID:G2636029; PIDN:CAB15566.1; PID:g2636075
A:Experimental source: strain 168
C:Comment: This protein has two functional regions: the amino-terminal region, which con
which carries the enhancing activity for these enzymes.
C:Genetics:
A:Gene: degU, iep
A:Start codon: GTG
C:Superfamily: regulatory protein comA; response regulator homology
C:Keywords: DNA binding; phosphoprotein; transcription regulation; two-component regulat
F:6-117/Domain: response regulator homology <RRH>
F:56/Binding site: phosphate (Asp) (covalent) #status predicted

Query Match 34.7%; Score 41; DB 1; Length 229;
Best Local Similarity 50.0%; Pred. No. 56;
Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 4 DSCEQQAQNVQHYH 17
| : : : : :
DB 36 DDGDEAARIVEHYH 49

RESULT 49
S21171
activin receptor STK9 - African clawed frog
C:Species: Xenopus laevis (African clawed frog)
C:Date: 22-Nov-1993 #sequence_revision 10-Nov-1995 #text_change 28-Feb-1997
C:Accession: S21171
R:Nishimatsu, S.; Oda, S.; Murakami, K.; Ueno, N.
FEBS Lett. 303, 81-84, 1992
A:Title: Multiple genes for Xenopus activin receptor expressed during early embryogenesis

A:Reference number: S21171; MUID:92275088
A:Accession: S21171
A:Molecule type: mRNA
A:Residues: 1-512 <NIS>
C:Superfamily: activin receptor II; protein kinase homology
C:Keywords: ATP
F:189-485/Domain: protein kinase homology <KIN>

Query Match 34.7%; Score 41; DB 2; Length 512;
Best Local Similarity 41.7%; Pred. No. 1.3e+02;
Matches 10; Conservative 2; Mismatches 8; Indels 4; Gaps 1;

QY 2 SPD-----SCEQQAQNVQHYHKNEL 21
| | | | | : : : : :
DB 97 SPDVFCCCEGACNACNERFYHSPEN 120
| | | | | : : : : :

RESULT 50
S39484
DNA-binding protein GT-2 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 07-Apr-1994 #sequence_revision 07-Apr-1994 #text_change 21-Jul-2000
C:Accession: S39484
R:Kuhn, R.M.; Caspar, T.; Dehesh, K.; Quail, P.H.
Plant Mol. Biol. 23, 337-348, 1993
A:Title: DNA binding factor GT-2 from Arabidopsis.
A:Reference number: S39484; MUID:94033312
A:Accession: S39484
A:Molecule type: mRNA
A:Residues: 1-575 <KUH>
A:Cross-references: EMBL:X72780; NID:g416489; PIDN:CAA51289.1; PID:g416490
C:Keywords: DNA binding

Query Match 34.7%; Score 41; DB 2; Length 575;
Best Local Similarity 41.2%; Pred. No. 1.5e+02;
Matches 7; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

QY 2 SPDSCQQAQNVQHYHK 18
| : : : : :
DB 86 SSKKCKEKFENVYKYHK 102

Search completed: December 17, 2001, 07:51:12
Job time: 94 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 17, 2001, 07:49:38 ; Search time 23.84 Seconds
(without alignments)
65.249 Million cell updates/sec

Title: US-09-462-625-2_COPY_55_75
Perfect score: 118
Sequence: 1 NSPSCQQQARNVQHYHKNEL 21

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 150 summaries

Database : A_Geneseq_1101.*

1: /SID88/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SID88/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SID88/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SID88/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SID88/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SID88/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SID88/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SID88/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SID88/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SID88/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SID88/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SID88/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SID88/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SID88/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SID88/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SID88/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SID88/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SID88/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SID88/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SID88/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SID88/gcgdata/geneseq/geneseq/AA2000.DAT.*
22: /SID88/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	118	100.0	181	18	Murine granulocyte
2	118	100.0	182	20	Mouse tag7 clone p
3	86	72.9	191	20	Human tag7 clone p
4	86	72.9	196	21	Human PRO1269 prot
5	86	72.9	196	21	Htag7 protein enco
6	86	72.9	196	21	Chondrosarcoma pep
7	86	72.9	196	21	Human PRO1269 (UNQ
8	86	72.9	196	22	protein of the inv
9	77	65.3	116	21	Human 5' EST relat
10	73	61.9	190	18	Bovine granulocyte
11	46	39.0	400	21	Human ORFX ORF2895

12	46	39.0	431	22	AA885400	Tumour endothelial
13	45.5	38.6	214	21	AA833039	Pinus radiata tran
14	45	38.1	184	19	AAW42098	Human Rab protein
15	44	37.3	33	22	AAW4285	Peptide #8322 enco
16	44	37.3	108	19	AAW74953	Human secreted pro
17	44	37.3	183	21	AA852083	Gene 27 human secr
18	44	37.3	183	21	AA852084	Human secreted pro
19	44	37.3	185	21	AA851559	Human RGD1 protein
20	44	37.3	186	19	AAW74804	Human secreted pro
21	44	37.3	213	22	AAE04179	Human gene 3 encod
22	44	37.3	427	20	AAV08218	Staphylococcus cap
23	43	36.4	162	21	AA840675	Human ORFX ORF439
24	43	36.4	384	22	AA892853	C glutamicum prote
25	43	36.4	388	22	AA879449	Corynebacterium gl
26	43	36.4	388	22	AA879450	Corynebacterium gl
27	43	36.4	443	18	AAW20743	H. pylori cytoplas
28	43	36.4	979	14	AA834783	Human CENP-C antiq
29	41.5	35.2	74	21	AA841011	Zea mays protein f
30	41.5	35.2	173	18	AAW20360	H. pylori secreted
31	41.5	35.2	198	18	AAW20963	H. pylori secreted
32	41	34.7	139	10	AA890358	Polypeptide with i
33	41	34.7	229	10	AA890995	B. subtilis sacUS2
34	41	34.7	229	10	AA891952	Peptide which prom
35	41	34.7	300	22	AA882403	S. epidermidis ope
36	41	34.7	459	19	AAW56738	Orpinomyces cellu
37	41	34.7	550	21	AAW07900	Arabidopsis thalia
38	41	34.7	575	21	AAW07899	Arabidopsis thalia
39	41	34.7	597	21	AAW07898	Arabidopsis thalia
40	41	34.7	716	20	AAW22176	Drosophila Acp36DE
41	40.5	34.3	327	21	AAW54095	Enzyme EPSJ involv
42	40.5	34.3	327	21	AAW43797	Amino acid sequenc
43	40	33.9	40	21	AA838431	Fragment of human
44	40	33.9	171	21	AA838428	Product of homeoti
45	40	33.9	231	14	AA843385	Arabidopsis thalia
46	40	33.9	246	21	AA859547	Arabidopsis thalia
47	40	33.9	250	21	AA859272	Arabidopsis thalia
48	40	33.9	250	21	AA859546	Arabidopsis thalia
49	40	33.9	262	21	AA819303	Arabidopsis thalia
50	40	33.9	262	21	AA850997	Arabidopsis thalia
51	40	33.9	312	21	AA859271	Arabidopsis thalia
52	40	33.9	312	21	AA859545	Arabidopsis thalia
53	40	33.9	319	22	AA890787	Human shear stress
54	40	33.9	331	21	AA858898	Breast and ovarian
55	40	33.9	336	21	AA819302	Arabidopsis thalia
56	40	33.9	336	21	AA850996	Arabidopsis thalia
57	40	33.9	359	21	AA819301	Arabidopsis thalia
58	40	33.9	359	21	AA850995	Arabidopsis thalia
59	40	33.9	361	21	AA859270	Arabidopsis thalia
60	40	33.9	392	22	AA881550	S. epidermidis ope
61	40	33.9	419	15	AA863440	Staphylococcus epi
62	40	33.9	463	21	AA843142	Human ORFX ORF2906
63	40	33.9	532	22	AA892514	C glutamicum prote
64	40	33.9	623	21	AAV54142	A human zinc finge
65	40	33.9	649	19	AAW53827	Pseudomonas XcpQ s
66	40	33.9	649	21	AA825594	Pseudomonas alcali
67	40	33.9	649	22	AA882556	Pseudomonas alcali
68	40	33.9	692	20	AA828653	Human Cytoplasmic
69	40	33.9	696	21	AA818181	Plasmodium falcipa
70	40	33.9	696	22	AA839775	Human polypeptide
71	40	33.9	807	20	AAW89247	Human PTP04. Homo
72	40	33.9	808	20	AA828652	Human Cytoplasmic
73	40	33.9	846	22	AAW41561	Human polypeptide
74	40	33.9	1221	22	AA891409	C glutamicum prote
75	40	33.9	1461	19	AAW64468	Human secreted pro
76	40	33.9	1461	22	AA890743	Human CW420-2 pro
77	39	33.1	20	21	AA867475	Tenebrio dessicati
78	39	33.1	97	22	AA835228	M tuberculosis Rv3
79	39	33.1	113	21	AA803200	Human secreted pro
80	39	33.1	120	21	AA870736	Human Wnt-4 protei
81	39	33.1	133	21	AA819788	Human Wnt-4 protei
82	39	33.1	173	19	AA837834	Recombinant peptid
83	39	33.1	173	19	AAW37836	Amino acid sequenc
84	39	33.1	196	19	AAW37837	Amino acid sequenc

85	39	33.1	196	19	AAW37835	Amino acid sequenc
86	39	33.1	227	20	AAI10844	Amino acid sequenc
87	39	33.1	253	20	AAW33724	Photobabidus lumen
88	39	33.1	276	19	AAW98845	H. pylori GHPO 158
89	39	33.1	279	22	AAW66503	Human ATP-dependen
90	39	33.1	281	22	AAW25780	Human protein sequ
91	39	33.1	304	21	AAW21677	Arabidopsis thalia
92	39	33.1	304	21	AAW51855	Arabidopsis thalia
93	39	33.1	351	19	AAW66276	Signal transductio
94	39	33.1	351	20	AAW411719	Human PRO864 (UNQ4
95	39	33.1	351	21	AAW44275	Human PRO864 (UNQ4
96	39	33.1	351	21	AAW57270	Wnt-4AF and Wnt-5c
97	39	33.1	351	21	AAW57272	Wnt-4AF and Wnt-5c
98	39	33.1	351	21	AAW57273	Wnt-4AF and Wnt-5c
99	39	33.1	351	22	AAW38889	Human polypeptide
100	39	33.1	365	22	AAW40675	Human polypeptide
101	39	33.1	619	21	AAW40554	Human ORFX ORF318
102	39	33.1	2048	22	AAW40027	Human polypeptide
103	39	33.1	2096	21	AAW41592	Human ORFX ORF1356
104	39	33.1	2139	22	AAW47278	PN7771. Homo sapi
105	38.5	32.6	378	21	AAW59149	Human GAK protein
106	38.5	32.6	1305	20	AAW79274	Protein kinase GAK
107	38	32.2	35	21	AAW88734	Core polypeptide f
108	38	32.2	35	22	AAW77089	Core polypeptide f
109	38	32.2	36	21	AAW88801	Core polypeptide f
110	38	32.2	99	21	AAW33666	Arabidopsis thalia
111	38	32.2	112	21	AAW23150	Human scaffold att
112	38	32.2	121	19	AAW86278	Blastx output of h
113	38	32.2	126	21	AAW03987	Human secreted pro
114	38	32.2	133	20	AAW59976	Human endometrium
115	38	32.2	133	21	AAW12184	Arabidopsis thalia
116	38	32.2	134	20	AAW36916	Protein which is s
117	38	32.2	136	21	AAW54311	Human pancreatic c
118	38	32.2	136	22	AAW37398	Human colon cancer
119	38	32.2	156	22	AAW21094	Peptide #7528 enco
120	38	32.2	163	22	AAW37293	Peptide #11330 enc
121	38	32.2	163	20	AAW48611	Human breast tumou
122	38	32.2	238	21	AAW56363	Human prostate can
123	38	32.2	240	21	AAW3721	Human cancer assoc
124	38	32.2	248	21	AAW53418	Human colon cancer
125	38	32.2	274	22	AAW02490	Arabidopsis thalia
126	38	32.2	275	20	AAW55952	Fission yeast SPE2
127	38	32.2	283	20	AAW31420	Mouse ath1j prote
128	38	32.2	283	21	AAW11972	Mouse ath1j prote
129	38	32.2	300	20	AAW31416	Human ASTH1J prote
130	38	32.2	300	21	AAW11968	Human ASTH1J prote
131	38	32.2	300	21	AAW92243	Human 36PiA6 trans
132	38	32.2	303	21	AAW54316	Human pancreatic c
133	38	32.2	303	22	AAW74438	Human colon cancer
134	38	32.2	304	22	AAW74283	Human colon cancer
135	38	32.2	313	21	AAW91284	Group B Streptococ
136	38	32.2	362	21	AAW42278	Human ORFX ORF2042
137	38	32.2	395	21	AAW91951	Human cytoskeleton
138	38	32.2	401	21	AAW42635	Human ORFX ORF2399
139	38	32.2	415	19	AAW69228	Human lysosomal si
140	38	32.2	418	20	AAW07475	Human ena/VASP-lik
141	38	32.2	418	21	AAW32828	Human ena/VASP-lik
142	38	32.2	431	20	AAW08217	Staphylococcus xyl
143	38	32.2	476	15	AAW47587	Rat bone formation
144	38	32.2	491	22	AAW40041	Human polypeptide
145	38	32.2	491	22	AAW59242	D.melanogaster cor
146	38	32.2	546	22	AAW92703	Human protein sequ
147	38	32.2	633	21	AAW94961	Human secreted pro
148	38	32.2	681	21	AAW84440	Amino acid sequenc
149	38	32.2	841	22	AAW59217	D.melanogaster cor
150	38	32.2	906	18	AAW25681	Transgenic mouse N

ALIGNMENTS

ID	AAW23723 standard; Protein; 181 AA.
XX	AAW23723;
AC	18-FEB-1998 (first entry)
XX	Murine granulocyte peptide A precursor (antimicrobial MGP-A).
DT	Antimicrobial peptide; antibiotic; antibacterial; antifungal;
DE	fungicide; antiprotozoa; protozoacide; antiviral; virucide;
XX	murine granulocyte peptide A; MGP-A; preservative; sepsis;
KW	endotoxaemia; mouse.
KW	Mus musculus.
XX	Key
XX	Location/Qualifiers
FT	Peptide
FT	/label= Mat_peptide
FT	/note= "MGP-A antimicrobial peptide (Claim 3)"
XX	W09729765-A1.
XX	21-AUG-1997.
XX	13-FEB-1997; 97WO-US02218.
XX	16-FEB-1996; 96US-0011834.
XX	(REGC) UNIV CALIFORNIA.
XX	Selsted ME;
XX	WPI; 1997-424753/39.
XX	N-PSDB; AAT78510.
XX	Antimicrobial (polypeptide(s) from bovine and murine granulocytes
XX	useful therapeutically, as preservatives for food, in water
XX	treatment and in agriculture
XX	Claim 9; Fig 5; 56pp; English.
XX	This protein comprises the precursor of a novel, claimed
XX	antimicrobial peptide from murine neutrophils, designated murine
XX	granulocyte peptide A or MGP-A (see AAW23725). Its amino acid
XX	sequence was deduced from a cDNA clone (see AAT78510) obtained from
XX	murine bone marrow. MGP-A and the bovine homologue, BGP-A (see
XX	AAW23724), exhibit activity against Gram-positive and Gram-negative
XX	bacteria, fungi and viruses, specifically Staphylococcus aureus,
XX	Escherichia coli, Candida albicans, Salmonella typhimurium and C.
XX	neoformans (claimed). They can be used in human or veterinary
XX	medicine (particularly to treat disorders associated with
XX	lipopolysaccharides, e.g. sepsis and endotoxaemia) or as
XX	preservatives in food products or in water supplies (claimed).
XX	They can also be applied to crops to reduce post-harvest spoilage
XX	or expressed in transgenic plants to increase their disease
XX	resistance. They have low immunogenicity. Carboxamidated analogues
XX	of MGP-A and BGP-A may also be used.
XX	Sequence 181 AA;
XX	Query Match 100.0%; Score 118; DB 18; Length 181;
XX	Best Local Similarity 100.0%; Pred. No. 1.6e-11;
XX	Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY	1 NSPDSCEQQQARNVQHYHKNEL 21
Db	
Db	55 nspdsceqqqarnvqhyhknel 75
RESULT 2	
AAW00770	
ID	AAW00770 standard; Protein; 182 AA.

PR 01-DEC-1999; 99WO-US28634.
 XX (GETH) GENENTECH INC.
 XX Botstein D, Goddard A, Gurney AL, Roy MA, Watanabe CK, Wood WI;
 XX N-PSDB; AAC58104.
 XX WPI; 2000-594320/56.
 XX N-PSDB; AAC58104.
 XX Antibodies specific for PRO polypeptides, used to diagnose and inhibit
 XX the growth of tumors in mammals, and to identify inhibitors of PRO
 XX polypeptide activity or expression
 XX Claim 61; Fig 4; 226pp; English.
 XX The present invention describes an antibody that binds to a human
 XX protein (I) selected from: PRO1269; PRO1410; PRO1755; PRO1780;
 XX PRO3434; PRO1297; PRO3567; PRO1293; PRO1303; PRO4344; PRO4354;
 XX PRO4397; PRO4407; PRO1555; PRO1096; PRO2038; and PRO2262. (I) has
 XX anticancer activity and can be used to diagnose tumours in mammals, by
 XX detecting complex formation when the antibody is contacted with test
 XX cells. Increased expression of genes encoding (I) can also be detected
 XX to diagnose tumours. Agents which inhibit the activity of (I),
 XX especially the antibodies, or an antisense oligonucleotide which
 XX hybridises to genes encoding (I), can be used to inhibit tumour growth,
 XX preferably by inducing cell death. Methods from the present invention
 XX can be used to identify compounds which inhibit the biological activity
 XX of (I). AAC58019 to AAC58102 represent PCR primers and hybridisation
 XX probes used in examples from the present invention for human PRO
 XX sequences. AAC58103 to AAC58122 and AAC24021 to AAC24040 represent human
 XX PRO polynucleotide and protein sequences given in the exemplification of
 XX the present invention.
 XX Sequence 196 AA;
 XX
 XX Query Match 72.9%; Score 86; DB 21; Length 196;
 XX Best Local Similarity 71.4%; Pred. No. 3e-06;
 XX Matches 15; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
 QY 1 NSPDCSQOQARNVQHYHKNEL 21
 DB 68 ntpascqqqgarnvqhymktl 88
 RESULT 5
 AAB25583
 ID AAB25583 standard; Protein; 196 AA.
 XX AAB25583;
 XX
 XX 21-NOV-2000 (first entry)
 XX Htag7 protein encoded by human secreted protein gene #8.
 XX
 XX Secreted protein; immunosuppressant; anti-inflammatory; antiarthritic;
 XX antirheumatic, dermatological; antiproliferative; antiarteriosclerotic;
 XX anticancer; vulnerary; antiviral; antibacterial; antifungal;
 XX immune disorder; Addison's disease; rheumatoid arthritis; dermatitis;
 XX multiple sclerosis; inflammatory disorder; inflammatory bowel disease;
 XX Crohn's disease; nephritis; hyperproliferative disorder;
 XX cardiovascular disorder; coronary arteriosclerosis; myocarditis; cancer;
 XX melanoma; lymphoma; wound healing; human.
 XX Homo sapiens.
 XX WO200029435-A1.
 XX 25-MAY-2000.
 XX 27-OCT-1999; 99WO-US25031.
 XX 28-OCT-1998; 98US-0105971.

XX (HUMA-) HUMAN GENOME SCI INC.
 XX Ni J, Ruben SM, Olsen HS, Young PE, Kenny JJ, Moore PA, Wei Y;
 XX Greene JM;
 XX WPI; 2000-387742/33.
 XX N-PSDB; AAA80613.
 XX Isolated nucleic acid molecules encoding human secreted proteins are
 XX used for the prevention, amelioration and treatment of autoimmune,
 XX inflammatory, hyperproliferative and cardiovascular disorders, cancer,
 XX wounds, and infectious diseases
 XX Claim 1; Figure 34; 803pp; English.
 XX The present invention relates to 12 secreted human proteins and the
 XX nucleotide sequences encoding them. The polynucleotide sequences given
 XX in AAA80606-AA80623 encode the 12 secreted protein sequences given in
 XX AAB25576-B25593. The human secreted proteins have various activities
 XX dependent on the tissues in which they are expressed. Examples of the
 XX activities of the proteins include: immunosuppressant;
 XX anti-inflammatory; antiarthritic; antirheumatic, dermatological;
 XX antiproliferative; antiarteriosclerotic; anticancer; vulnerary;
 XX antiviral; antibacterial; and antifungal activity. The proteins,
 XX polypeptides, agonists and antagonists may be used to treat prevent
 XX and/or diagnose various disease, disorders and conditions examples of
 XX which include: immune disorders e.g. Addison's disease, rheumatoid
 XX arthritis, dermatitis, and multiple sclerosis; inflammatory disorders
 XX e.g. inflammatory bowel disease, Crohn's disease and nephritis;
 XX hyperproliferative disorders such as paraproteinemia and purpura;
 XX cardiovascular disorders e.g. coronary arteriosclerosis and myocarditis;
 XX cancer e.g. melanoma and lymphoma. The proteins and polynucleotide
 XX sequences may also be used in wound healing and the treatment of
 XX infectious diseases. The human secreted protein gene #8 and protein
 XX sequences are represented in sequences AAA80613 and AAB25583. Sequences
 XX AAA80662-AA80663 represent genes related to the secreted protein gene#8.
 XX Sequence 196 AA;
 XX
 XX Query Match 72.9%; Score 86; DB 21; Length 196;
 XX Best Local Similarity 71.4%; Pred. No. 3e-06;
 XX Matches 15; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
 QY 1 NSPDCSQOQARNVQHYHKNEL 21
 DB 68 ntpascqqqgarnvqhymktl 88
 RESULT 6
 AAY96964
 ID AAY96964 standard; Protein; 196 AA.
 XX AAY96964;
 XX
 XX 31-OCT-2000 (first entry)
 XX Chondrosarcoma peptidoglycan recognition protein-like protein.
 XX
 XX Peptidoglycan recognition protein-like protein; PGRP; PGRP-C; regulator;
 XX chondrosarcoma; immunosuppressive; antibacterial; anti-apoptotic;
 XX tumour necrosis factor; TNF; anti-infectious; mononuclear phagocyte;
 XX inhibitor; protein co-ordinate data.
 XX Homo sapiens.
 XX OS
 XX Key Location/Qualifiers
 XX Peptide 1..21
 XX /label= Signal_peptide
 XX Protein 22..196
 XX /label= Mature_protein
 XX Domain 34..117

FT Region /label= PGRP-like_domain
FT 1..6
FT /label= Antigenic_region
FT 20..29
FT /label= Antigenic_region
FT 33..43
FT /label= Antigenic_region
FT 63..79
FT /label= Antigenic_region
FT 99..112
FT /label= Antigenic_region
FT 133..146
FT /label= Antigenic_region
FT 160..165
FT /label= Antigenic_region
FT 168..181
FT /label= Antigenic_region
FT 190..196
FT /label= Antigenic_region

XX WO200039327-A1.

PN 06-JUL-2000.

XX 22-DEC-1999; 99WO-US30736.

XX 23-DEC-1998; 98US-0113809.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Rosen CA, Ruben SM, Young PE, Olsen HS;

XX WPI; 2000-452414/39.

DR N-PSDB; AAA51719.

XX Polynucleotide encoding peptidoglycan recognition protein-like protein,
PT antibodies specific to it useful for preventing, treating conditions
PT e.g. endotoxic shock and auto-immune disorders and infections in mammal

XX Claim 11; Fig 3; 19lpp; English.

XX Novel human peptidoglycan recognition protein-like proteins (PGRP)
CC expressed by keratinocytes (PGRP-K), wound healing tissues (PGRP-W)
CC or chondrosarcoma tissue (PGRP-C) are disclosed. The PGRP DNA and
CC proteins are useful for preventing, treating or ameliorating a medical
CC condition in a mammal (claimed). PGRP is useful in augmenting the immune
CC system in such areas as immune recognition, antigen presentation and
CC immune system activation. Antibodies or antagonists directed against
CC these proteins may be useful in reducing or eliminating disorders
CC associated with tumour necrosis factor (TNF) and TNF-like cytokines,
CC such as endotoxic shock and autoimmune disorders and for treating
CC infectious diseases including silicosis, sarcoidosis and idiopathic
CC pulmonary fibrosis.

XX Sequence 196 AA;

Query Match 72.9%; Score 86; DB 21; Length 196;
Best Local Similarity 71.4%; Pred. No. 3e-06;
Matches 15; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Oy 1 NSPSCSQOARVQHYHKNEL 21
Db 68 ntpascqgqarnvghymk1 88

RESULT 7

AAAY99400
ID AAAY99400 standard; Protein; 196 AA.

XX AC AAAY99400;

XX DT 08-AUG-2000 (first entry)

XX Human PRO1269 (UNQ639) amino acid sequence SEQ ID NO:216.
DE Human; PRO polypeptide; membrane bound protein; receptor; diagnosis;
XX transmembrane; secretion; immunoadhesion; pharmaceutical; screening.
KW Homo sapiens.
OS
XX WO200012708-A2.
PN 09-MAR-2000.
XX 01-SEP-1999; 99WO-US20111.
XX 01-SEP-1998; 98US-0098716.
XX 01-SEP-1998; 98US-0098749.
XX 01-SEP-1998; 98US-0098750.
XX 02-SEP-1998; 98US-0098803.
XX 02-SEP-1998; 98US-0098821.
XX 02-SEP-1998; 98US-0098843.
XX 02-SEP-1998; 98US-0099536.
XX 09-SEP-1998; 98US-0099596.
XX 09-SEP-1998; 98US-0099598.
XX 09-SEP-1998; 98US-0099602.
XX 09-SEP-1998; 98US-0099642.
XX 10-SEP-1998; 98US-0099741.
XX 10-SEP-1998; 98US-0099754.
XX 10-SEP-1998; 98US-0099763.
XX 10-SEP-1998; 98US-0099792.
XX 10-SEP-1998; 98US-0099808.
XX 10-SEP-1998; 98US-0099812.
XX 10-SEP-1998; 98US-0099815.
XX 10-SEP-1998; 98US-0099816.
XX 15-SEP-1998; 98US-0100385.
XX 15-SEP-1998; 98US-0100388.
XX 15-SEP-1998; 98US-0100390.
XX 16-SEP-1998; 98US-0100584.
XX 16-SEP-1998; 98US-0100627.
XX 16-SEP-1998; 98US-0100661.
XX 16-SEP-1998; 98US-0100662.
XX 16-SEP-1998; 98US-0100664.
XX 17-SEP-1998; 98US-0100683.
XX 17-SEP-1998; 98US-0100684.
XX 17-SEP-1998; 98US-0100710.
XX 17-SEP-1998; 98US-0100711.
XX 17-SEP-1998; 98US-0100919.
XX 17-SEP-1998; 98US-0100930.
XX 18-SEP-1998; 98US-0100848.
XX 18-SEP-1998; 98US-0100849.
XX 18-SEP-1998; 98US-0101014.
XX 18-SEP-1998; 98US-0101068.
XX 18-SEP-1998; 98US-0101071.
XX 22-SEP-1998; 98US-0101279.
XX 23-SEP-1998; 98US-0101471.
XX 23-SEP-1998; 98US-0101472.
XX 23-SEP-1998; 98US-0101474.
XX 23-SEP-1998; 98US-0101475.
XX 23-SEP-1998; 98US-0101476.
XX 23-SEP-1998; 98US-0101477.
XX 23-SEP-1998; 98US-0101479.
XX 24-SEP-1998; 98US-0101738.
XX 24-SEP-1998; 98US-0101741.
XX 24-SEP-1998; 98US-0101743.
XX 24-SEP-1998; 98US-0101915.
XX 24-SEP-1998; 98US-0101916.
XX 24-SEP-1998; 98US-0102207.
XX 29-SEP-1998; 98US-0102240.
XX 29-SEP-1998; 98US-0102307.
XX 29-SEP-1998; 98US-0102330.
XX 29-SEP-1998; 98US-0102331.
XX 30-SEP-1998; 98US-0102484.
XX 30-SEP-1998; 98US-0102487.
XX 30-SEP-1998; 98US-0102570.

small molecule inhibitors of the relevant receptor/ligand interactions
 Claim 12; Fig 122; 773pp; English.

AAA37022 to AAA37144 encode the new isolated human transmembrane, receptor or secreted PRO polypeptides given in AAY99340 to AAY99462. The transmembrane and receptor PRO proteins can be used for screening of potential peptide or small molecule inhibitors of the relevant receptor/ligand interactions. The polypeptides and nucleotide sequences encoding them have various industrial applications, including uses as pharmaceutical and diagnostic agents. AAA37145 to AAA37330 represent PCR primers and hybridisation probes used in the isolation of the PRO polypeptides from the present invention.

Sequence 196 AA;

Query Match 72.9%; Score 86; DB 21; Length 196;

Best Local Similarity 71.4%; Pred. NO. 3e-06; 4; Indels 0; Gaps 0; Matches 15; Conservative 2; Mismatches 0;

QY 1 NSPDSCEQQAQNVQVHKNEL 21

Db 68 ntpscqgqgarnvghymklt 88

RESULT 8

AAB66149 AAB66149 standard; protein; 196 AA.

XX AAB66149;

AC 02-APR-2001 (first entry)

DT Protein of the invention #61.

DE Secreted; transmembrane; gene therapy.

XX Unidentified.

XX WO200078961-A1.

XX 28-DEC-2000.

XX 18-FEB-2000; 2000WO-US04342.

XX 23-JUN-1999; 99US-0141037.

XX 20-JUL-1999; 99US-0144758.

XX 26-JUL-1999; 99US-0145698.

XX 01-SEP-1999; 99WO-US20111.

XX 29-OCT-1999; 99US-0162506.

XX 30-NOV-1999; 99WO-US28313.

XX 02-DEC-1999; 99WO-US28551.

XX 16-DEC-1999; 99WO-US30095.

XX 05-JAN-2000; 2000WO-US00219.

XX 06-JAN-2000; 2000WO-US00376.

(GETH) GENENTECH INC.

PA Baker KP, Botstein D, Desnoyers L, Eaton DL, Ferrara N, Fong S;

PI Gao W, Goddard A, Godowski PJ, Grimaldi CJ, Gurney AL, Hillan KJ;

PI Pan J, Paoni NF, Roy MA, Smith V, Stewart TA, Tumas D;

PI Watanabe CK, Williams FW, Wood WI;

XX WPI; 2001-071395/08.

XX Secreted and transmembrane proteins and nucleic acids designated PRO,

PT useful as hybridization probes, in chromosome and gene mapping and gene

PT therapy -

XX Claim 1; Fig 122; 787pp; English.

XX The present invention relates to secreted and transmembrane proteins.

XX CC

30-SEP-1998; 98US-0102571.
 01-OCT-1998; 98US-0102684.
 01-OCT-1998; 98US-0102687.
 02-OCT-1998; 98US-0102965.
 06-OCT-1998; 98US-0103258.
 06-OCT-1998; 98US-0103449.
 07-OCT-1998; 98US-0103314.
 07-OCT-1998; 98US-0103315.
 07-OCT-1998; 98US-0103328.
 07-OCT-1998; 98US-0103395.
 07-OCT-1998; 98US-0103396.
 07-OCT-1998; 98US-0103401.
 08-OCT-1998; 98US-0103678.
 08-OCT-1998; 98US-0103679.
 08-OCT-1998; 98US-0103711.
 14-OCT-1998; 98US-0104257.
 20-OCT-1998; 98US-0104987.
 20-OCT-1998; 98US-0105000.
 20-OCT-1998; 98US-0105002.
 21-OCT-1998; 98US-0105104.
 22-OCT-1998; 98US-0105169.
 22-OCT-1998; 98US-0105266.
 26-OCT-1998; 98US-0105693.
 26-OCT-1998; 98US-0105694.
 27-OCT-1998; 98US-0105807.
 27-OCT-1998; 98US-0105881.
 27-OCT-1998; 98US-0105882.
 27-OCT-1998; 98US-0106062.
 28-OCT-1998; 98US-0106023.
 28-OCT-1998; 98US-0106030.
 28-OCT-1998; 98US-0106032.
 28-OCT-1998; 98US-0106033.
 28-OCT-1998; 98US-0106178.
 28-OCT-1998; 98US-0106248.
 29-OCT-1998; 98US-0106384.
 29-OCT-1998; 98US-0108500.
 30-OCT-1998; 98US-0106464.
 03-NOV-1998; 98US-0106856.
 03-NOV-1998; 98US-0106902.
 03-NOV-1998; 98US-0106905.
 03-NOV-1998; 98US-0106919.
 03-NOV-1998; 98US-0106932.
 03-NOV-1998; 98US-0106934.
 10-NOV-1998; 98US-0107783.
 17-NOV-1998; 98US-0108775.
 17-NOV-1998; 98US-0108779.
 17-NOV-1998; 98US-0108787.
 17-NOV-1998; 98US-0108788.
 17-NOV-1998; 98US-0108801.
 17-NOV-1998; 98US-0108802.
 17-NOV-1998; 98US-0108806.
 17-NOV-1998; 98US-0108807.
 17-NOV-1998; 98US-0108867.
 17-NOV-1998; 98US-0108925.
 18-NOV-1998; 98US-0108948.
 18-NOV-1998; 98US-0108849.
 18-NOV-1998; 98US-0108850.
 18-NOV-1998; 98US-0108851.
 18-NOV-1998; 98US-0108852.
 18-NOV-1998; 98US-0108858.
 18-NOV-1998; 98US-0108904.

(GETH) GENENTECH INC.

Baker K, Goddard A, Gurney AL, Smith V, Watanabe CK, Wood WI;

WPI; 2000-237871/20.

N-PSDB; AAA37082.

New mammalian DNA sequences encoding transmembrane, receptor or secreted PRO polypeptides, useful for screening of potential peptide or

CC These proteins and the DNA encoding them may be used as hybridization
 CC probes, in chromosome and gene mapping and in the generation of
 CC anti-sense RNA and DNA. They may also be used to generate either
 CC transgenic animals or knockout animals which are in turn useful for
 CC development and screening of therapeutically useful reagents.
 CC The nucleic acids may also be used in gene therapy.
 XX
 SQ Sequence 196 AA;

Query Match 72.9%; Score 86; DB 22; Length 196;
 Best Local Similarity 71.4%; Pred. NO. 3e-06;
 Matches 15; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 NSPDSCEQOARNVQHYHKNEL 21
 I: ||:|||||||
 Db 68 ntpascqgqarinvqhyhmktl 88

RESULT 9
 AAY64935
 ID AAY64935 standard; Protein; 116 AA.
 XX
 AC AAY64935;

XX 01-FEB-2000 (first entry)
 XX
 DE Human 5' EST related polypeptide SEQ ID NO:1096.

XX Human; 5' EST; expressed sequence tag; secreted protein; diagnosis;
 KW gene therapy; chromosome mapping; upstream regulatory sequence;
 KW forensic; location; development; protein synthesis; stability;
 KW regulation; identification.
 XX
 OS Homo sapiens.

XX WO953051-A2.
 XX 21-OCT-1999.
 XX 09-APR-1999; 99WO-IB00712.
 XX 09-APR-1998; 98US-0057719.
 XX 28-APR-1998; 98US-0069047.
 XX (GEST) GENSET.
 XX Dumas Milne Edwards J, Duclert A, Giordano J;
 PI WPI; 2000-038446/03.
 XX N-PSDB; AA242549.

PT Novel secreted protein 5' expressed sequence tag sequences used in
 PT diagnostic, forensic, gene therapy, and chromosome mapping procedures
 XX
 PS Claim 3; Page 685; 837pp; English.

XX AA242265 to AA243075 represent novel 5' expressed sequence tag (EST)
 CC sequences, corresponding to human secreted proteins. AAY64651 to
 CC AAY6438 represent the EST-related proteins corresponding to AA242265 to
 CC AA243052. The 5' ESTs can be used for producing secreted human gene
 CC products. They can be used to identify and isolate 5' untranslated
 CC regions (UTRs) and upstream regulatory regions which control the
 CC location, development stage, rate, and quantity of protein synthesis, as
 CC well as stability of mRNA. The ESTs are also useful as probes for
 CC chromosome mapping, and to obtain full length cDNA clones. The ESTs can
 CC also be used in forensic procedures to identify individuals, or in
 CC diagnostic procedures to identify individuals having genetic diseases
 CC resulting from abnormal gene expression. The products may also be used in
 CC gene therapy protocols. The nucleic acids encoding signal peptides can be
 CC used for directing extracellular secretion of a polypeptide or the
 CC insertion of a polypeptide into a membrane, or importing a polypeptide
 CC into a cell. The proteins encoded by the EST sequences may be useful in

CC treating a variety of human conditions. Secreted proteins have
 CC therapeutic value, and the identification of new secreted proteins is
 CC valuable. AA242249 to AA242264 and AAY64644 to AAY64650 represent
 CC sequences used in the exemplification of the present invention.
 XX
 SQ Sequence 116 AA;

Query Match 65.3%; Score 77; DB 21; Length 116;
 Best Local Similarity 66.7%; Pred. NO. 4.8e-05;
 Matches 14; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 1 NSPDSCEQOARNVQHYHKNEL 21
 I: ||:|||||||
 Db 68 ntxascqgqarinvqhyhmktl 88

RESULT 10
 AAW23722
 ID AAW23722 standard; Protein; 190 AA.
 XX
 AC AAW23722;

XX 18-FEB-1998 (first entry)

XX Bovine granulocyte peptide A precursor (antimicrobial BGP-A).
 XX Antimicrobial peptide; antibiotic; antibacterial; antifungal;
 KW fungicide; antiprotozoa; protozoacide; antiviral; virucide;
 KW bovine granulocyte peptide A; BGP-A; preservative; sepsis;
 KW endotoxaemia; cattle.
 XX
 OS Bos taurus.

XX Key Location/Qualifiers
 FH Peptide 1..21
 FT /label= Sig_peptide
 FT 22..177
 FT /label= Propeptide
 FT 178..190
 FT /label= Mat_peptide
 FT /note= "BGP-A antimicrobial peptide (Claim 2)"

XX WO9729765-A1.
 XX 21-AUG-1997.
 XX 13-FEB-1997; 97WO-US02218.
 XX 16-FEB-1996; 96US-0011834.
 XX (REGC) UNIV CALIFORNIA.
 XX Selsted ME;
 XX WPI; 1997-424753/39.
 XX N-PSDB; AAT78509.

XX Antimicrobial (poly)peptide(s) from bovine and murine granulocytes -
 PT useful therapeutically, as preservatives for food, in water
 PT treatment and in agriculture
 XX
 PS Claim 8; Fig 4; 56pp; English.

XX This protein comprises the precursor (see AAW23722) of a novel,
 CC claimed antimicrobial peptide from bovine neutrophils, designated
 CC bovine granulocyte peptide A or BGP-A (see AAW23724). Its amino acid
 CC sequence was deduced from a cDNA clone (see AAT78509) obtained from
 CC bovine bone marrow. BGP-A and the murine homologue, MGP-A (see
 CC AAW23725), exhibit activity against Gram-positive and Gram-negative
 CC bacteria, fungi and viruses, specifically Staphylococcus aureus,
 CC Escherichia coli, Candida albicans, Salmonella typhimurium and C.
 CC neoformans (claimed). They can be used in human or veterinary

XX AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397, CC CC which represent the human ORFX open reading frames 1 to 3161. The ORFX CC

XX
CC
The invention provides novel human stem cell growth factor-like

sequences have activities such as: cytostatic; hepatotropic; vulnary; antiparasitic; antiparkinsonian; nootropic; immunoprotective; osteopathic; anticonvulsant; antiarthritic; immunosuppressant; immunostimulant; cardiant; thermolytic; coagulant; vasotropic; antidiabetic; hypotensive; dermatological; immunosuppressive; antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic; antithyroid; and antianemic. The sequences can be used for determining the presence of or predisposition to, or preventing or treating pathological conditions associated with an ORFX-associated disorder. The nucleic acids can be used to express ORFX proteins in gene therapy vectors. The proteins and nucleic acids may be used to treat cancers, proliferative disorders, neurodegenerative disorders, osteoarthritis, graft vs host disease, cardiovascular disease, diabetes mellitus, hypertension, hypothyroidism, cholesterol ester storage, systemic lupus erythematosus, severe combined immunodeficiency (SCID), AIDS, viral bacterial or fungal infection, malaria, autoimmune disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and cartilage damage, nocturnal haemoglobinuria, antiinflammatory disease; to enhance nucleation; to inhibit thrombin; and as a contraceptive.

The present invention relates to novel plant transcription factors from *Eucalyptus grandis* or *Pinus radiata*. The present sequence is one such transcription factor. The transcription factor may be used to produce a plant having modified gene expression such as a woody plant e.g. a *eucalyptus*, pine, acacia, poplar, sweetgum, teak, or mahogany species or to modify the activity of a polypeptide in a plant. The transcription factors of the present invention are members from the following families of regulatory proteins: bZIP, bZIP family of G-box binding factors, basic

Wed Dec 19 08:45:36 2001

us-09-462-625-2_copy_55_75.rag

```
QY      4 DSCEQOQARNVOHYHKN 19
Db      61 escasavrnveyyqsn 76

RESULT 15
AAM34285
ID      AAM34285 standard; Protein; 33 AA.
XX      AC
XX      AAM34285;
XX      17-OCT-2001 (first entry)
XX      DE
XX      Peptide #8322 encoded by probe for measuring placental gene expression.
XX      KW
XX      Probe; microarray; human; placenta; antenatal diagnosis;
XX      genetic disorder.
XX      OS
XX      Homo sapiens.
XX      PN
XX      WQ200157272-A2.
XX      09-AUG-2001.
XX      30-JAN-2001; 2001WO-US00663.
XX      04-FEB-2000; 2000US-0180312.
XX      26-MAY-2000; 2000US-0207456.
XX      30-JUN-2000; 2000US-0608408.
XX      03-AUG-2000; 2000US-0632366.
XX      21-SEP-2000; 2000US-0234687.
XX      27-SEP-2000; 2000US-0236359.
XX      04-OCT-2000; 2000GB-0024263.
XX      PA
XX      (MOLE-) MOLECULAR DYNAMICS INC.
XX      PI
XX      Penn SG, Hanzel DK, Chen W, Rank DR;
XX      WPI; 2001-488897/53.
XX      Human genome-derived single exon nucleic acid probes useful for
XX      analyzing gene expression in human placenta -
XX      Claim 27; SEQ ID NO 34554; 654pp; English.
XX      The present invention relates to single exon nucleic acid probes (SENP:
XX      see AA1315-AA157546). The present sequence is a peptide encoded by one
XX      such probe. The probes are useful for producing a microarray for
XX      predicting, measuring and displaying gene expression in samples derived
XX      from human placenta. The probes are useful for antenatal diagnosis of
XX      human genetic disorders.
XX      SQ
XX      Sequence 33 AA;

Query Match 37.3%; Score 44; DB 22; Length 33;
Best Local Similarity 40.0%; Pred. No. 2.8;
Matches 8; Conservative 4; Mismatches 8; Indels 0; Gaps 0;

QY      2 SPDSCEQOQARNVOHYHKNEL 21
Db      3 apqpcqgahpvdrrwrev 22

RESULT 16
AAM74953
ID      AAM74953 standard; Protein; 108 AA.
XX      AC
XX      AAM74953;
XX      25-JAN-1999 (first entry)
XX      DE
XX      Human secreted protein encoded by gene 75 clone HB1AB39.
XX      KW
XX      Human; secreted protein; testis; tumour; foetal brain tissue;
XX      fusion protein; cancer; central nervous system; seizure;
XX      diagnosis; neurodegenerative disease.
XX      OS
XX      Homo sapiens.
XX      FH
XX      Key
XX      Misc-difference 108
XX      Location/Qualifiers
XX      /label= unknown
XX      WO9839448-A2.
XX      11-SEP-1998.
XX      06-MAR-1998; 98WO-US04493.
XX      02-OCT-1997; 97US-0061060.
XX      07-MAR-1997; 97US-0038621.
XX      07-MAR-1997; 97US-0040161.
XX      07-MAR-1997; 97US-0040162.
XX      07-MAR-1997; 97US-0040163.
XX      07-MAR-1997; 97US-0040333.
XX      07-MAR-1997; 97US-0040334.
XX      07-MAR-1997; 97US-0040336.
XX      07-MAR-1997; 97US-0040626.
XX      11-APR-1997; 97US-0043311.
XX      11-APR-1997; 97US-0043312.
XX      11-APR-1997; 97US-0043313.
XX      11-APR-1997; 97US-0043314.
XX      11-APR-1997; 97US-0043568.
XX      11-APR-1997; 97US-0043569.
XX      11-APR-1997; 97US-0043576.
XX      11-APR-1997; 97US-0043578.
XX      11-APR-1997; 97US-0043580.
XX      11-APR-1997; 97US-0043669.
XX      11-APR-1997; 97US-0043670.
XX      11-APR-1997; 97US-0043671.
XX      11-APR-1997; 97US-0043672.
XX      11-APR-1997; 97US-0043674.
XX      23-MAY-1997; 97US-0047492.
XX      23-MAY-1997; 97US-0047500.
XX      23-MAY-1997; 97US-0047501.
XX      23-MAY-1997; 97US-0047502.
XX      23-MAY-1997; 97US-0047503.
XX      23-MAY-1997; 97US-0047581.
XX      23-MAY-1997; 97US-0047582.
XX      23-MAY-1997; 97US-0047583.
XX      23-MAY-1997; 97US-0047584.
XX      23-MAY-1997; 97US-0047585.
XX      23-MAY-1997; 97US-0047586.
XX      23-MAY-1997; 97US-0047587.
XX      23-MAY-1997; 97US-0047588.
XX      23-MAY-1997; 97US-0047589.
XX      23-MAY-1997; 97US-0047590.
XX      23-MAY-1997; 97US-0047592.
XX      23-MAY-1997; 97US-0047593.
XX      23-MAY-1997; 97US-0047594.
XX      23-MAY-1997; 97US-0047595.
XX      23-MAY-1997; 97US-0047596.
XX      23-MAY-1997; 97US-0047597.
XX      23-MAY-1997; 97US-0047598.
XX      23-MAY-1997; 97US-0047599.
XX      23-MAY-1997; 97US-0047600.
XX      23-MAY-1997; 97US-0047601.
XX      23-MAY-1997; 97US-0047612.
XX      23-MAY-1997; 97US-0047613.
XX      23-MAY-1997; 97US-0047614.
XX      23-MAY-1997; 97US-0047615.
XX      23-MAY-1997; 97US-0047617.
XX      23-MAY-1997; 97US-0047618.
XX      23-MAY-1997; 97US-0047632.
```


CC myocardial ischaemias; wound healing; neurological diseases such as
CC cerebral anoxia and epilepsy; and infectious diseases such as viral,
CC bacterial, fungal and parasitic infections may also be treated using the
CC proteins and polynucleotides of the invention. Sequences AAC95512 -
CC AAC95520 and AAB52011 are used in the isolation and characterisation of
CC the proteins and polynucleotides of the invention.
XX
SQ Sequence 183 AA;

Query Match 37.3%; Score 44; DB 21; Length 183;
Best Local Similarity 36.8%; Pred. No. 20;
Matches 7; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

QY 1 NSPDSCEQQARNVQHYHKN 19
I : I:: I::: I
Db 57 nvglcqlrvrnveyqgsn 75

RESULT 18
AAB52084
ID AAB52084 standard; Protein: 183 AA.
XX
AC AAB52084;
XX
DT 21-FEB-2001 (first entry)
XX
DE Human secreted protein sequence encoded by gene 27 SEQ ID NO:133.
XX
KW Human; secreted protein; cytostatic; immunosuppressive; nootropic;
KW neuroprotective; antiviral; antiallergic; hepatotropic; antidiabetic;
KW antiinflammatory; antiulcer; vulnerary; anticonvulsant; antibacterial;
KW antifungal; antiparasitic; cardiant; cancer; immune disease; allergy;
KW cardiovascular disorder; wound healing; infection; neurological disease.
XX
OS Homo sapiens.
XX
PN WO2000061596-A1.
XX
PD 19-OCT-2000.
XX
PF 06-APR-2000; 2000WO-US08983.
XX
PR 09-APR-1999; 99US-0128703.
PR 20-JAN-2000; 2000US-0176068.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
PA (ROSE/) ROSEN C A.
XX
XX Ruben SM, Komatsoulis G;
XX
XX WPI; 2000-611865/58.
XX N-PSDB; AAC95547.
XX
XX Fifty nucleic acid molecules encoding human secreted proteins, useful
XX in the prevention, treatment and diagnosis of cancer, immune disorders,
XX cardiovascular disorders and neurological diseases -
XX
XX Disclosure; Page 50; 505pp; English.
XX
XX Polynucleotide sequences AAC95521 - AAC95570 represent cDNA encoding 50
XX human secreted proteins AAB52012 - AAB52161. Sequences AAB52062 -
XX AAB52103 represent alternative polypeptides encoded by the genes, and
XX amino acid sequences with which they share homology. The genes and
XX proteins have activities dependent on the tissues and cells in which they
XX are expressed. Examples of their activities include cytostatic;
XX immunosuppressive; nootropic; neuroprotective; antiviral; antiallergic;
XX hepatotropic; antidiabetic; antiinflammatory; antiulcer; vulnerary;
XX anticonvulsant; antibacterial; antifungal; antiparasitic; and cardiant.
XX The secreted proteins, polypeptides, and agonists may be
XX useful in treating, preventing and/or diagnosing diseases and disorders
XX such as cancer, particularly breast and ovarian cancer, and other cancers
XX of the adrenal gland, bone, bone marrow, breast, gastrointestinal tract,

CC liver, lung, or urogenital. Immune disorders such as Addison's disease,
CC allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,
CC diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid
CC arthritis and ulcerative colitis; cardiovascular disorders such as
CC myocardial ischaemias; wound healing; neurological diseases such as
CC cerebral anoxia and epilepsy; and infectious diseases such as viral,
CC bacterial, fungal and parasitic infections may also be treated using the
CC proteins and polynucleotides of the invention. Sequences AAC95512 -
CC AAC95520 and AAB52011 are used in the isolation and characterisation of
XX the proteins and polynucleotides of the invention.
SQ Sequence 183 AA;

Query Match 37.3%; Score 44; DB 21; Length 183;
Best Local Similarity 36.8%; Pred. No. 20;
Matches 7; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

QY 1 NSPDSCEQQARNVQHYHKN 19
I : I:: I::: I
Db 57 nlglcqlrvrnveyqgsn 75

RESULT 19
AAY51559
ID AAY51559 standard; Protein: 185 AA.
XX
AC AAY51559;
XX
DT 18-MAY-2000 (first entry)
XX
DE Human RGD1 protein.
XX
KW RGD1; Rab-specific guanine-nucleotide dissociation inhibitor; anti-HIV;
KW antineurodegenerative; anti-ischemic; antianemic; vasotropic; detection;
KW hepatotropic; cytostatic; anti-inflammatory; gene therapy; apoptosis;
KW acquired immune deficiency syndrome; neurodegeneration; aplastic anemia;
KW retinitis pigmentosa; ischemic injury; cirrhosis; cancer; inflammation;
KW allergy; Crohn's disease; multiple sclerosis; diagnosis.
XX
OS Homo sapiens.
XX
XX US6015672-A.
XX
PD 18-JAN-2000.
XX
PF 03-DEC-1997; 97US-0984295.
XX
PR 03-DEC-1997; 97US-0984295.
XX
PA (INCY-) INCYTE PHARM INC.
XX
XX Yue H, Lal P, Corley NC;
PI
XX WPI; 2000-181140/16.
DR N-PSDB; AA288761.
XX
XX New human nucleic acid encoding guanine-nucleotide dissociation
XX inhibitor, used for diagnosis, treatment and prevention of abnormal
XX apoptosis such as neurodegeneration or cancer -
XX
XX Claim 1; Column 35-36; 25pp; English.
XX
XX This invention describes a novel isolated and purified polynucleotide
XX sequence (I) of 802 base pairs encoding the human Rab-specific
XX guanine-nucleotide dissociation inhibitor (RGDI) polypeptide (II). The
XX protein described in the method of the invention has anti-HIV;
XX antineurodegenerative, anti-ischemic, antianemic, vasotropic,
XX hepatotropic, cytostatic and anti-inflammatory activity. (I) is used for
XX recombinant production of RGDI, in gene therapy, including expression of
XX complements of (I), as a source of antisense, triplex-forming or ribozyme
XX therapeutics, as a source of probes and primers for diagnosis and
XX monitoring of RGDI-related diseases, in usual amplification and/or

CC hybridization tests, for genomic mapping and for detecting mutations.
CC RGD1 is used to treat or prevent disorders associated with increased
CC apoptosis, e.g. acquired immune deficiency syndrome, neurodegeneration,
CC retinitis pigmentosa, aplastic anemia, ischemic injuries, and cirrhosis.
CC RGD1 is used to raise specific antibodies (Ab) or to screen for specific
CC antagonists and agonists which are potential therapeutic agents.
CC Antagonists of RGD1 are used to treat abnormal cellular proliferation,
CC e.g. many forms of cancer and inflammation such as allergy, Crohn's
CC disease and multiple sclerosis. Ab are used as therapeutic antagonists
CC (directly or for targeted delivery of drugs), to diagnose and monitor
CC RGD1-related diseases, in competitive drug-screens and to isolate RGD1
CC from natural sources. This sequence represents the human RGD1 protein
CC described in the method of the invention.
XX Sequence 185 AA;
SQ

Query Match 37.3%; Score 44; DB 21; Length 185;
Best Local Similarity 36.8%; Pred. No. 20;
Matches 7; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

QY 1 NSPDSCEQQRNVQHYHKN 19
| : | : | | : | |
Db 59 nlgelcqlrvnveyyqsn 77

RESULT 20

AAW74804
ID AAW74804 standard; Protein; 186 AA.

AC AAW74804;

XX 25-JAN-1999 (first entry)

DE Human secreted protein encoded by gene 75 clone HB1AB39.

XX Human; secreted protein; testis; tumour; foetal brain tissue;
KW fusion protein; cancer; central nervous system; seizure;
KW diagnosis; neurodegenerative disease.

OS Homo sapiens.

XX Key Location/Qualifiers
FH Misc-difference 186
FT label= unknown

XX WO9839448-A2.

XX 11-SEP-1998.

XX 06-MAR-1998; 98WO-US04493.

XX 02-OCT-1997; 97US-0061060.

PR 07-MAR-1997; 97US-0038621.

PR 07-MAR-1997; 97US-0040161.

PR 07-MAR-1997; 97US-0040162.

PR 07-MAR-1997; 97US-0040163.

PR 07-MAR-1997; 97US-0040333.

PR 07-MAR-1997; 97US-0040334.

PR 07-MAR-1997; 97US-0040336.

PR 07-MAR-1997; 97US-0040626.

PR 11-APR-1997; 97US-0043311.

PR 11-APR-1997; 97US-0043312.

PR 11-APR-1997; 97US-0043313.

PR 11-APR-1997; 97US-0043314.

PR 11-APR-1997; 97US-0043672.
PR 11-APR-1997; 97US-0043674.
PR 23-MAY-1997; 97US-0047492.
PR 23-MAY-1997; 97US-0047500.
PR 23-MAY-1997; 97US-0047501.
PR 23-MAY-1997; 97US-0047502.
PR 23-MAY-1997; 97US-0047503.
PR 23-MAY-1997; 97US-0047581.
PR 23-MAY-1997; 97US-0047582.
PR 23-MAY-1997; 97US-0047583.
PR 23-MAY-1997; 97US-0047584.
PR 23-MAY-1997; 97US-0047585.
PR 23-MAY-1997; 97US-0047586.
PR 23-MAY-1997; 97US-0047587.
PR 23-MAY-1997; 97US-0047588.
PR 23-MAY-1997; 97US-0047589.
PR 23-MAY-1997; 97US-0047590.
PR 23-MAY-1997; 97US-0047592.
PR 23-MAY-1997; 97US-0047593.
PR 23-MAY-1997; 97US-0047594.
PR 23-MAY-1997; 97US-0047595.
PR 23-MAY-1997; 97US-0047596.
PR 23-MAY-1997; 97US-0047597.
PR 23-MAY-1997; 97US-0047598.
PR 23-MAY-1997; 97US-0047599.
PR 23-MAY-1997; 97US-0047600.
PR 23-MAY-1997; 97US-0047601.
PR 23-MAY-1997; 97US-0047612.
PR 23-MAY-1997; 97US-0047613.
PR 23-MAY-1997; 97US-0047614.
PR 23-MAY-1997; 97US-0047615.
PR 23-MAY-1997; 97US-0047617.
PR 23-MAY-1997; 97US-0047618.
PR 23-MAY-1997; 97US-0047632.
PR 23-MAY-1997; 97US-0047633.
PR 06-JUN-1997; 97US-0048964.
PR 06-JUN-1997; 97US-0048974.
PR 13-JUN-1997; 97US-0049610.
PR 16-JUL-1997; 97US-0051926.
PR 18-AUG-1997; 97US-0052874.
PR 22-AUG-1997; 97US-0055724.
PR 22-AUG-1997; 97US-0056630.
PR 22-AUG-1997; 97US-0056631.
PR 22-AUG-1997; 97US-0056632.
PR 22-AUG-1997; 97US-0056636.
PR 22-AUG-1997; 97US-0056637.
PR 22-AUG-1997; 97US-0056662.
PR 22-AUG-1997; 97US-0056664.
PR 22-AUG-1997; 97US-0056845.
PR 22-AUG-1997; 97US-0056862.
PR 22-AUG-1997; 97US-0056864.
PR 22-AUG-1997; 97US-0056872.
PR 22-AUG-1997; 97US-0056874.
PR 22-AUG-1997; 97US-0056875.
PR 22-AUG-1997; 97US-0056876.
PR 22-AUG-1997; 97US-0056877.
PR 22-AUG-1997; 97US-0056878.
PR 22-AUG-1997; 97US-0056879.
PR 22-AUG-1997; 97US-0056880.
PR 22-AUG-1997; 97US-0056881.
PR 22-AUG-1997; 97US-0056882.
PR 22-AUG-1997; 97US-0056884.
PR 22-AUG-1997; 97US-0056886.
PR 22-AUG-1997; 97US-0056887.
PR 22-AUG-1997; 97US-0056888.
PR 22-AUG-1997; 97US-0056889.
PR 22-AUG-1997; 97US-0056892.
PR 22-AUG-1997; 97US-0056893.
PR 22-AUG-1997; 97US-0056894.
PR 22-AUG-1997; 97US-0056903.
PR 22-AUG-1997; 97US-0056908.
PR 22-AUG-1997; 97US-0056909.
PR 22-AUG-1997; 97US-0056910.

XX FemA; identification; detection; therapy; infection; femB;
KW amplification; genotyping; gram-positive bacteria; vaccine.
XX
OS *Staphylococcus capitis*.

	key	Location/Qualifiers
PH	Misc-difference 412	
FT		/note= "in frame stop codon"
FT		

PN WO9916780-A2.

08-APR-1999.

28-SEP-1998:

XX
PP 26-SEP-1997. 97EP-0870146XX
PA (BENAL) BELGIAN MIN NAT DEEE

PA (UYLO-) UNIV CATHOLIQUE LOUVAIN.

PI Gala J, Vannuffel P;

DR WPI; 1999-287521/24.

XX
XX

PS Claim 22: Fig 9a-b; 48pp; English.

This invention describes novel *Staphylococcus*-specific oligonucleotides based on the consensus femA nucleotide sequence which are used to develop products for the identification, detection and therapy of infections. The oligonucleotides can be used for the genetic amplification, the identification and/or quantification of various femA sequences which are specific to known or unknown *Staphylococci* species. Since the femA sequence is similar to the femB sequence, the oligonucleotides can also be used for the molecular genotyping of femB genes of different *Staphylococci* species or other gram-positive bacteria. The femA nucleic acids can also be used in therapeutic applications. They can also be used to identify inhibitors, e.g. antibodies or antisense oligonucleotides, for blocking expression of the femA nucleotide sequences. They can also be used for producing vaccines against *Staphylococci* infections.

AA	Sequence	427 AA;
SQ		

Query Match	37.3%	Score 44	DB 20	Length 427
Best Local Similarity	35.5%	Pred. No. 53		
Matches 11	Conservative	4	Mismatches 6	Indels 10
				Gaps 17

Qy 1 NSPDSCEQ-----ARNVQHYHKNEL 21
| | : : ||
pb 275 nkrdnlggaldangqkideaknlqgheqnel 305

RESULT 23

AAB40673
ID AAB40675 standard: Protein: 162 AA.

XX
AC
AAB40675.

XX
DE 08-FEB-2007 . /first entry)XX
XX
XX

"-----" CMMY CMY420 -el-junt+;do cccm:ccc CEO ID NO-979

XX	Human; open reading frame; ORF _X ; detection;
KW	cytostatic; hepatotropic;
KW	vulnery; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
KW	anticonvulsant; osteopathic; antichrictic; immunosuppressant; cardiac;
KW	immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW	hypotensive; dermatological; immunosuppressive; antinflammatory;
KW	antiviral; antibacterial; antifungal; antirheumatic; antithyroid;

antidote; gene therapy; cancer; proliferative disorder; hypertension; neurodegenerative disorder; osteoarthritis; graft vs host disease; cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS; cholesterol ester storage; systemic lupus erythematosus; infection; severe combined immunodeficiency; malaria; autoimmune disorder; asthma; allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound; bone damage; cartilage damage; antiinflammatory disease; coagulation; thrombosis; contraceptive.

OS Homo sapiens.

AA : WO200058473-A2.

XX
05-OCT-2000XX
DE
, 31-MAR-2000:XX
21 MAR 1960
0045-0137607

PR 02-APR-1999; 99US-0127636.
PR 05-APR-1999; 99US-0127728

PR 30-MAR-2000; 2000US-0540763.

PA (CURA-) CURAGEN CORP.

PI Shimkets RA, Leach M;

DR WPI; 2000-602362/57.

XX
XX

PT NOVEL nuclear acids
PT useful for treating

XX
P.T.
neutroneutron

PS Claim II; Page 900; 330/pp; English: vv

CC AAC74446 to AAC7606 encode the proteins given in AAB40237 to AAB40238, which represent the human ORFX open reading frames 1 to 3161. The

sequences have activities such as: cytostatic; hepatocarcinogenic; tumorigenic; immunosuppressive; antiproliferative; antiparkinsonian; nootropic; neuroprotective; antipsoriatic; antiparkinsonian; antihistaminic; immunosuppressive; osteopathic; anticonvulsant; antihistaminic; coagulant; vasotropic; immunostimulant; cardiac; thrombolytic; coagulant; vasotropic; antiadrenergic; hypotensive; dermatological; immunosuppressive; antiflamatory; antibacterial; antiviral; antifungal; antirheumatic; antithyroid; and antianemic. The sequences can be used for determining pathological conditions associated with an ORFX-associated disorder. The presence of or predisposition to, or preventing or treating pathological conditions associated with an ORFX proteins in gene therapy. Nucleic acids can be used to express ORFX proteins in gene therapy. The proteins and nucleic acids may be used to treat cancers, proliferative disorders, neurodegenerative disorders, osteoarthritis, graft vs host disease, cardiovascular disease, diabetes mellitus, hypertension, hypothyroidism, cholesterol ester storage, systemic lupus erythematosus, severe combined immunodeficiency (SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and cartilage damage, nocturnal haemoglobinuria, antiinflammatory disease; to enhance coagulation; to inhibit thrombosis; and as a contraceptive.

AA	Sequence	162 AA;
SO		

Query Match 36.4%; Score 43; DB 21; Length 162;
Best Local Similarity 42.9%; Pred. No. 25;
Matches 6; Conservative 6; Mismatches 2; Indels

QY . 6 CEQQARNVQHYNK 19
||::|::|::|:
dh 146 ceeagavrevhad 159

RESULT 24

AAG2653
 ID AAG2653 standard; Protein; 384 AA.
 XX

C	AAG92653;
X	26-SEP-2001 (first entry)
T	C glutamicum protein fragment SEQ ID NO: 6407.
X	Coryneform bacterium; amino acid synthesis; vitamin; saccharide; organic acid synthesis.
E	Corynebacterium glutamicum.
X	EPI108790-A2.
X	20-JUN-2001.
X	18-DEC-2000; 2000EP-0127688.
F	16-DEC-1999; 99JP-0377484.
R	07-APR-2000; 2000JP-0159162.
R	03-AUG-2000; 2000JP-0280988.
X	(KYOW) KYOWA HAKKO KOGYO KK.
A	Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H; Tateishi N, Senoh A, Ikeda M, Ozaki A;
I	WPI: 2001-376931/40.
I	N-PSDB: AAH67872.
DR	Novel polynucleotides derived from Coryneform bacteria, for identifying mutation point of a gene, measuring expression of a gene, analysing expression profile or pattern of a gene and identifying homologous gene
PT	Claim 17: SEQ ID NO: 6407; 246pp + Sequence Listing; English.
XX	The present invention provides a number of nucleotide and protein sequences from the Coryneform bacterium Corynebacterium glutamicum. These are useful for identifying the mutation point of a gene derived from a mutant of coryneform bacterium, measuring expression amount and analysing the expression profile or expression pattern of a gene derived from Coryneform bacterium, and identifying a homologue of a gene derived from coryneform bacterium. Coryneform bacteria are useful for producing amino acids, nucleic acids, vitamins, saccharides and organic acids, particularly L-lysine. The present sequence is a protein described in the exemplification of the invention. CC Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from the European Patent Office.
CC	Sequence 384 AA;
SQ	Query Match 36.4%; Score 43; DB 22; Length 384; Best Local Similarity 43.8%; Pred No. 69; Matches 7; Conservative 2; Mismatches 7; Indels 0; Gaps 0;
QY	3 PDSEQQQAARVQHKK 18 I : IIII II:
Db	75 ptgtqgarnaayhr 90
RESULT 25	
AAAB79449	
ID	AAAB79449 standard; Protein; 388 AA.
AC	AAAB79449;
XX	
DT	30-APR-2001 (first entry)
XX	Corynebacterium glutamicum SMP protein sequence SEQ ID NO:414.
DE	Corynebacterium glutamicum; carbon metabolism and energy production;
XX	

CC C. glutamicum, identify and localise C. glutamicum sequences of interest,
CC in evolutionary studies, in determining SMP protein regions required
CC for function, in modulating SMP protein activity, in modulating the
CC metabolism of sugars, and in modulating high-energy molecule production
CC in a cell (i.e. ATP, NADPH).
XX
SQ Sequence 388 AA;

Query Match 36.4%; Score 43; DB 22; Length 388;
Best Local Similarity 43.8%; Pred. No. 69;
Matches 7; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 3 PDSCQQQARNVQHYHK 18
Db 79 ptgtgdqarnaayhr 94
:
: ||||| ||:

RESULT 26
AAB79450
ID AAB79450 standard; Protein; 388 AA.
XX
AC AAB79450;
XX
DT 30-APR-2001 (first entry)
XX
DE Corynebacterium glutamicum SMP protein sequence SEQ ID NO:416.
XX
KW Corynebacterium glutamicum; carbon metabolism and energy production;
KW SMP protein; sugar metabolism and oxidative phosphorylation protein;
KW fine chemical production; organic acid; proteinogenic amino acid;
KW nonproteinogenic amino acid; purine base; pyrimidine base; nucleoside;
KW nucleotide; lipid; saturated fatty acid; unsaturated fatty acid; diol;
KW carbohydrate; aromatic compound; vitamin; cofactor; polyketide; enzyme;
KW diagnosis; Corynebacterium diphtheriae; evolutionary study.
XX
OS Corynebacterium glutamicum.
XX
PN WO200100844-A2.
XX
PD 04-JAN-2001.
XX
PF 23-JUN-2000; 2000WO-IB00943.
XX
PR 25-JUN-1999; 99US-0141031.
PR 08-JUL-1999; 99DE-1031412.
PR 08-JUL-1999; 99DE-1031413.
PR 08-JUL-1999; 99DE-1031419.
PR 08-JUL-1999; 99DE-1031420.
PR 08-JUL-1999; 99DE-1031424.
PR 08-JUL-1999; 99DE-1031428.
PR 08-JUL-1999; 99DE-1031431.
PR 08-JUL-1999; 99DE-1031433.
PR 08-JUL-1999; 99DE-1031434.
PR 08-JUL-1999; 99DE-1031510.
PR 08-JUL-1999; 99DE-1031562.
PR 08-JUL-1999; 99DE-1031634.
PR 09-JUL-1999; 99DE-1032180.
PR 09-JUL-1999; 99DE-1032227.
PR 09-JUL-1999; 99DE-1032230.
PR 09-JUL-1999; 99DE-1032230.
PR 09-JUL-1999; 99US-0143208.
PR 14-JUL-1999; 99DE-1032924.
PR 14-JUL-1999; 99DE-1032973.
PR 14-JUL-1999; 99DE-1033005.
PR 27-AUG-1999; 99DE-1040765.
PR 31-AUG-1999; 99US-0151572.
PR 03-SEP-1999; 99DE-1042076.
PR 03-SEP-1999; 99DE-1042079.
PR 03-SEP-1999; 99DE-1042086.
PR 03-SEP-1999; 99DE-1042087.
PR 03-SEP-1999; 99DE-1042088.
PR 03-SEP-1999; 99DE-1042095.
PR 03-SEP-1999; 99DE-1042123.

PR 03-SEP-1999; 99DE-1042125.
XX
PA (BADI) BASF AG.
XX
PI Pompejus M, Kroeger B, Schroeder H, Zelder O, Haberhauer G;
XX
DR WPI; 2001-061975/07.
DR N-ESDB; AAF71567.
XX
XX New isolated Corynebacterium glutamicum nucleic acid encoding a sugar
PT metabolism and oxidative phosphorylation protein for production or
PT modulation of production of fine chemicals e.g. amino acids,
PT carbohydrates or enzymes
XX
PS Claim 20; Page 741-742; 1246pp; English.
XX
CC AAF71360 to AAF71750 encode the Corynebacterium glutamicum sugar
CC metabolism and oxidative phosphorylation (SMP) proteins given in
CC AAB79243 to AAB 79633 which are involved in carbon metabolism and
CC energy production. The C. glutamicum SMP gene can be used in vectors
CC (II) for expression in host cells and production or modulation of
CC production of fine chemicals, such as, an organic acid, a proteinogenic
CC or nonproteinogenic amino acid (preferred), a purine or pyrimidine base,
CC a nucleoside, a nucleotide, a lipid, a saturated or unsaturated fatty
CC acid, a diol, a carbohydrate, an aromatic compound, a vitamin, a
CC cofactor, a polyketide, or an enzyme. The presence of (I) or SMP proteins
CC (III) encoded by them are used for diagnosing the presence or activity of
CC Corynebacterium diphtheriae in a subject. (I), (II), (III) or host cells
CC containing them are used to map genomes of organisms related to
CC C. glutamicum, identify and localise C. glutamicum sequences of interest,
CC in evolutionary studies, in determining SMP protein regions required
CC for function, in modulating SMP protein activity, in modulating the
CC metabolism of sugars, and in modulating high-energy molecule production
CC in a cell (i.e. ATP, NADPH).
XX
SQ Sequence 388 AA;

Query Match 36.4%; Score 43; DB 22; Length 388;
Best Local Similarity 43.8%; Pred. No. 69;
Matches 7; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 3 PDSCQQQARNVQHYHK 18
Db 79 ptgtgdqarnaayhr 94
:
: ||||| ||:

RESULT 27
AAW20743
ID AAW20743 standard; protein; 443 AA.
XX
AC AAW20743;
XX
DT 16-JUL-1997 (first entry)
XX
DE H. Pylori cytoplasmic protein, 06eel0709orf5.
XX
KW Cytoplasmic; vaccine; prevention; treatment; infection; identification;
KW binding compound; bacterium; life cycle; activator; bacteria; inhibitor;
KW duodenal ulcer disease; chronic gastritis; diagnosis; envelope.
XX
OS Helicobacter pylori.
XX
PN WO9640893-A1.
XX
PD 19-DEC-1996.
XX
PF 06-JUN-1996; 96WO-US09122.
XX
PR 01-APR-1996; 96US-0630405.
PR 07-JUN-1995; 95US-0487032.
XX
PA (ASTR) ASTRA AB.

Wed Dec 19 08:45:36 2001

XX Berglindh OT, Smith D, Mellgaard BL;
 XX WPI; 1997-052306/05.
 DR N-PSDB; AAT67996.
 XX Helicobacter pylori nucleic acid sequences and related
 PT polypeptide(s) - useful for vaccines to treat or prevent H. pylori
 PT infection, and to detect Helicobacter
 XX
 PS Claim 61; Page 1158-1159; 1481pp; English.
 XX The present sequence is a Helicobacter pylori cytoplasmic protein
 CC involved in cofactor metabolism.
 CC The protein may be used in a vaccine to prevent or treat H. pylori
 CC infection or to identify H. pylori polypeptide binding compounds,
 CC useful as potential H. pylori life cycle activators or inhibitors.
 CC The genomic sequence of H. pylori (ATCC 55679) was determined from
 CC overlapping contigs generated by mechanically shearing the bacterial
 CC DNA. The sequences were analysed for ORF of at least 180 nucleotides,
 CC and the predicted coding regions defined by computer evaluation. To
 CC identify likely H. pylori antigens for vaccine development, the amino
 CC acid sequences predicted from various ORF were analysed for significant
 CC homology to other known or exported membrane proteins. Having identified
 CC and determined the sequences of interest, particular regions can be
 CC isolated from H. pylori by PCR amplification for recombinant polypeptide
 CC production, e.g. in E. coli hosts.
 XX
 SQ Sequence 443 AA;

Query Match 36.4%; Score 43; DB 18; Length 443;
 Best Local Similarity 41.2%; Pred. No. 81;
 Matches 7; Conservative 6; Mismatches 4; Indels 0; Gaps 0;
 QY 5 SCEQOARNVQHYNEL 21
 Db 235 scekiavnrhlqrsev 251

RESULT 28
 AAR34783
 ID AAR34783 standard; Protein; 979 AA.
 XX
 AC AAR34783;
 XX
 DT 14-JUL-1993 (first entry)
 XX
 DE Human CENP-C antigen.
 XX
 KW Autoimmune rheumatic diseases; CREST syndrome; classification;
 KW diagnosis; immunoassay.
 XX
 OS Homo sapiens.
 XX
 PN WO9305151-A.
 XX
 PD 18-MAR-1993.
 XX
 PF 11-SEP-1992; 92WO-US07649.
 XX
 PR 12-SEP-1991; 91US-0759001.
 XX
 PA (UYJO) UNIV JOHNS HOPKINS.
 XX
 PI Earnshaw WC, Saltooh H, Tomkiel JE;
 XX
 DR WPI; 1993-100976/12.
 DR N-PSDB; AAQ38525.
 XX
 PT DNA encoding human CENP-C and its antibodies - for detecting
 PT anti-centromere antibodies in patients with auto-immune diseases
 XX

Claim 17; Page 45; 68pp; English.

PS The sequence is that of the human CENP-C antigen which may be
 XX recombinantly produced in large amounts. Thus the polypeptide is
 CC available to more accurately classify patients with such auto-
 CC immune rheumatic diseases as CREST syndrome, and commercially
 CC useful quantities are available for use in diagnostic systems.
 XX
 SQ Sequence 979 AA;

Query Match 36.4%; Score 43; DB 14; Length 979;
 Best Local Similarity 41.2%; Pred. No. 2e+02;
 Matches 7; Conservative 5; Mismatches 5; Indels 0; Gaps 0;
 QY 1 NSPDSCEQOARNVQHYH 17
 Db 165 ntpdskkissrindhh 181

RESULT 29
 AAG41011
 ID AAG41011 standard; Protein; 74 AA.
 XX
 AC AAG41011;
 XX
 DT 18-OCT-2000 (first entry)
 XX
 DE Zea mays protein fragment SEQ ID NO: 50968.

Protein identification; signal transduction pathway; metabolic pathway;
 KW hybridisation assay; genetic mapping; gene expression control; promoter;
 KW termination sequence; corn.
 XX
 OS Zea mays subsp. mays.

EP1033405-A2.
 PD 06-SEP-2000.
 XX
 PF 25-FEB-2000; 2000EP-0301439.
 XX
 PR 25-FEB-1999; 99US-0121825.
 PR 05-MAR-1999; 99US-0123180.
 PR 09-MAR-1999; 99US-0123548.
 PR 23-MAR-1999; 99US-0125788.
 PR 25-MAR-1999; 99US-0126264.
 PR 29-MAR-1999; 99US-0126785.
 PR 01-APR-1999; 99US-0127462.
 PR 06-APR-1999; 99US-0128234.
 PR 08-APR-1999; 99US-0128714.
 PR 16-APR-1999; 99US-0129845.
 PR 19-APR-1999; 99US-0130077.
 PR 21-APR-1999; 99US-0130449.
 PR 23-APR-1999; 99US-0130510.
 PR 28-APR-1999; 99US-0131449.
 PR 30-APR-1999; 99US-0132048.
 PR 04-MAY-1999; 99US-0132407.
 PR 05-MAY-1999; 99US-0132484.
 PR 06-MAY-1999; 99US-0132485.
 PR 06-MAY-1999; 99US-0132486.
 PR 07-MAY-1999; 99US-0132487.
 PR 11-MAY-1999; 99US-0132863.
 PR 14-MAY-1999; 99US-0134219.
 PR 14-MAY-1999; 99US-0134221.
 PR 14-MAY-1999; 99US-0134370.
 PR 18-MAY-1999; 99US-0134768.
 PR 19-MAY-1999; 99US-0134941.
 PR 21-MAY-1999; 99US-0135124.
 PR 21-MAY-1999; 99US-0135353.

PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142820.
PR 12-JUL-1999; 99US-0142877.
PR 13-JUL-1999; 99US-0143342.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 10-SEP-1999; 99US-0152363.
PR 13-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 22-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.

QY 1 NSPDS-CEQQARNVOHYHNE 20
 Db 76 ntpnsgckvndlnfkhyhnie 96

Query Match 35.2%; Score 41.5; DB 18; Length 198;
 Best Local Similarity 42.9%; Pred. No. 56;
 Matches 9; Conservative 4; Mismatches 7; Indels 1; Gaps 1;

QY 1 NSPDS-CEQQARNVOHYHNE 20
 Db 124 ntpnsgckvndlnfkhyhnie 144

RESULT 32

QY 28-OCT-1999; 99US-0161982.
 PR 28-OCT-1999; 99US-0161993.
 PR 29-OCT-1999; 99US-0162142.

Query Match 35.2%; Score 41.5; DB 21; Length 74;
 Best Local Similarity 66.7%; Pred. No. 18;
 Matches 10; Conservative 0; Mismatches 4; Indels 1; Gaps 1;

QY 3 PDSCE-QOARNVOHY 16
 Db 28 porceaqarleqy 42

RESULT 30

AAW20560
 ID AAW20560 standard; Protein; 173 AA.

XX AC AAW20560;
 DT 15-JUL-1997 (first entry)
 DE H. pylori secreted or periplasmic protein 663530.aa.
 KW Cytoplasmic; vaccine; prevention; treatment; infection; identification;
 KW binding compound; bacterium; life cycle; activator; inhibitor;
 KW duodenal ulcer disease; chronic gastritis; diagnosis; envelope.
 OS Helicobacter pylori.
 XX WO9640893-AL.
 PD 19-DEC-1996.
 PF 06-JUN-1996; 96WO-US09122.
 PR 01-APR-1996; 96US-0630405.
 PR 07-JUN-1995; 95US-0487032.
 XX (ASTR) ASTRA AB.
 PI Berglindh OT, Smith D, Mellgaerd BL;
 DR WPI; 1997-052306/05.
 DR N-PSDB; AAT67707.

XX Helicobacter pylori nucleic acid sequences and related
 polypeptide(s) - useful for vaccines to treat or prevent H. pylori
 infection, and to detect Helicobacter

Claim 72; Page 715; 1481pp; English.

This sequence is a H. pylori secreted or periplasmic protein.
 The protein may be used in a vaccine to prevent or treat H. pylori
 infection or to identify H. pylori polypeptide binding compounds,
 useful as potential H. pylori life cycle activators or inhibitors.
 The genomic sequence of H. pylori (ATCC 55679) was determined from
 overlapping contigs generated by mechanically shearing the bacterial
 DNA. The sequences were analysed for ORF of at least 180 nucleotides,
 and the predicted coding regions defined by computer evaluation. To
 identify likely H. pylori antigens for vaccine development, the amino
 acid sequences predicted from various ORF were analysed for significant
 homology to other known or exported membrane proteins. Having identified
 and determined the sequences of interest, particular regions can be
 isolated from H. pylori by PCR amplification for recombinant polypeptide
 production, e.g. in E. coli hosts.

XX SQ Sequence 173 AA;

Query Match 35.2%; Score 41.5; DB 18; Length 173;
 Best Local Similarity 42.9%; Pred. No. 48;
 Matches 9; Conservative 4; Mismatches 7; Indels 1; Gaps 1;

```

AAP90358
ID AAP90358 standard; protein; 139 AA.
XX
AC AAP90358;
XX
DT 01-NOV-1989 (first entry)
XX
DE Polypeptide with inhibitory effect on prodn. of extracellular
DE protease of Bacillus subtilis.
XX
KW peptide; inhibitor; Bacillus subtilis; extracellular protease.
XX
OS Bacillus subtilis.
XX
PN JP01124386-A.
XX
PD 17-MAY-1989.
XX
PF 09-NOV-1987; 87JP-0282382.
XX
PR 09-NOV-1987; 87JP-0282382.
XX
PA (MITU ) MITSUBISHI KASEI CORP.
XX
DR WPI; 1989-187408/26.
DR N-PSDB; AAN90122.
XX
PT DNA fragment encoding specific amino acid sequence - has inhibitory
PT activity for extracellular protease of Bacillus subtilis.
XX
PS Claim 1; page 14; 15pp; Japanese.
XX
CC The peptide has an inhibitory effect on the prodn. of extracellular
CC protease in Bacillus subtilis. See also AAN90122.
XX
SQ Sequence 139 AA;

Query Match 34.7%; Score 41; DB 10; Length 139;
Best Local Similarity 50.0%; Pred. No. 45;
Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 4 DSCQQQARNVQHYH 17
Db 36 dggdeaarivehvh 49

RESULT 33
AAP90995
ID AAP90995 standard; protein; 229 AA.
XX
AC AAP90995;
XX
DT 23-FEB-1990 (first entry)
XX
DE B. subtilis sacUS2 polypeptide.
XX
KW Levan saccharase
XX
OS Bacillus subtilis.
XX
PN W08909264-A.
XX
PR 22-MAR-1989; 89WO-FR00134.
XX
PD 05-OCT-1989.
XX
PF 22-MAR-1988; 88FR-0003736.
XX
PA (INSP ) INST PASTEUR.
XX
PI Kunst F, Debarbouille M, Msadek T, Rapoport G, Klier A, Dedonder R;
ID AAG82403 standard; Protein; 300 AA.

WPI; 1989-309530/42.
DR N-PSDB; AAN91619.
XX
DNA contg. Bacillus subtilis sacU locus - for inducing overproduction
in microorganisms
XX
Claim 3; Fig 6; 87pp; French.
XX
SacUS2 confers DegU phenotype, and is encoded by part of the 2.55 kb
Sali-Sphi SacU locus. This can restore levan saccharase synthesis in
B. subtilis sacU- mutants, and overproduces proteins in this or other
microorganisms.
XX
SQ Sequence 229 AA;

Query Match 34.7%; Score 41; DB 10; Length 229;
Best Local Similarity 50.0%; Pred. No. 80;
Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 4 DSCQQQARNVQHYH 17
Db 36 dggdeaarivehvh 49

RESULT 34
AAP91952
ID AAP91952 standard; peptide; 229 AA.
XX
AC AAP91952;
XX
DT 12-FEB-1990 (first entry)
XX
DE Peptide which promotes formn. of B. subtilis extracellular protease.
XX
KW Bacillus subtilis; extracellular protease.
XX
PN JP01218588-A.
XX
PD 31-AUG-1989.
XX
PF 25-FEB-1988; 88JP-0040815.
XX
PR 25-FEB-1988; 88JP-0040815.
XX
PA (MITU ) MITSUBISHI KASEI CORP.
XX
DR WPI; 1989-296609/41.
DR N-PSDB; AAN90668.
XX
PT DNA fragment encoding specific amino acid sequence - promotes formation
PT of Bacillus subtilis extracellular protease.
XX
PS Claim 1; page 439; 16pp; Japanese.
XX
CC The peptide has an ability to promote formn. of B.subtilis extracellular
CC protease. It is expressed in B.subtilis in high yield.
XX
SQ Sequence 229 AA;

Query Match 34.7%; Score 41; DB 10; Length 229;
Best Local Similarity 50.0%; Pred. No. 80;
Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 4 DSCQQQARNVQHYH 17
Db 36 dggdeaarivehvh 49

RESULT 35
AAG82403
ID AAG82403 standard; Protein; 300 AA.

```

XX AAG82403;
AC
XX
XX
DT
XX
XX
DE 03-SEP-2001 (first entry)
XX
XX S. epidermidis open reading frame protein sequence SEQ ID NO:1900.
XX
XX Staphylococcus epidermidis SR1 strain; infection; diagnosis;
KW vaccination; endocarditis.
XX
XX Staphylococcus epidermidis.
OS
XX W0200134809-A2.
PN
XX
XX 17-MAY-2001.
PD
XX
XX 09-NOV-2000; 2000WO-US30782.
PF
XX
XX 09-NOV-1999; 99US-0164258.
PR
XX (GLAX) GLAXO GROUP LTD.
PA
XX Kimmerly WJ;
PI
XX
XX WPI; 2001-316495/33.
DR N-PSDB; AAH53253.
XX
XX Nucleic acids encoding polypeptides from Staphylococcus epidermidis,
XX useful for vaccinating against infections, e.g. endocarditis -
PT
XX
XX Claim 18; Page 519; 2188pp; English.
PS
XX AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides
CC (II), given in AAG81454 to AAG83120, from Staphylococcus epidermidis.
CC (I) and (II) can have antibacterial activity and therefore can be used
CC in vaccination. The nucleic acids (I) may be used to produce the
CC S. epidermidis polypeptides (II) via the production of vectors
CC containing them which are used to produce hosts cells which express the
CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be
CC used to vaccinate subjects and to raise antibodies against the bacteria.
CC The polypeptides may also be used to assay for other inhibitors of their
CC activity and therefore identify compounds that may be used for the
CC treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to
CC AAH55090 represent specifically claimed S. epidermidis genomic DNA
CC polynucleotide sequences from the present invention. AAH55091 to
CC AAH55098 represent oligonucleotide sequences and primers which are used
CC in the exemplification of the present invention.
CC N.B. The present invention specifically claims all the polynucleotide
CC sequences given in the sequence listing of the present specification,
CC however the sequence listing only goes up to SEQ ID NO:4454 so even
CC though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,
CC no sequences are present for SEQ ID NO:4455 to 4464.
XX
XX Sequence 300 AA;
SQ

Query Match 34.7%; Score 41; DB 22; Length 300;
Best Local Similarity 70.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 11 RNVOHYHKN 20
DB 184 rdelhyhkn 193

RESULT 36
AAW56738
ID AAW56738 standard; Protein; 459 AA.
XX
XX AC AAW56738;
XX
XX DT 14-SEP-1998 (first entry)
XX

DE Orpinomyces cellulase celsA.
XX
XX Cellulase; endoglucanase; cellobiohydrolase; celsA.
XX
XX OS Orpinomyces sp. strain PC-2.
XX
XX PN W09814597-A1.
XX
XX PD 09-APR-1998.
XX
XX PF 03-OCT-1997; 97WO-US18008.
XX
XX PR 04-OCT-1996; 96US-0027883.
XX
XX PA (UYGE-) UNIV GEORGIA RES FOUND INC.
XX
XX PI Chen H, Li X, Ljungdahl LG;
XX
XX DR WPI; 1998-240096/21.
XX N-PSDB; AAV29471.
XX
XX PT New recombinant DNA encoding Orpinomyces cellulase protein - useful
XX for, e.g. producing recombinant Orpinomyces cellulase in host cell
XX
XX PS Claim 1; Page 28-30; 69pp; English.
XX
XX CC This polypeptide comprises cellulase celsA of Orpinomyces sp. strain
CC OC-2, an anaerobic bovine rumen fungus Orpinomyces sp. PC-2. Its
CC amino acid sequence was deduced from an isolated cDNA clone (see
CC AAV29471). CelsA has endoglucanase and cellobiohydrolase activity,
CC with highest activity at pH 4.8 and 50 degC. CelsA (see AAW56742)
CC and celsC (see AAW56739) cellulolytic enzymes of Orpinomyces sp. PC-2
CC are also provided. Recombinant DNA molecules encoding Orpinomyces
CC cellulase proteins are claimed, as well as recombinant cells
CC selected from Saccharomyces cerevisiae, Escherichia coli,
CC Aspergillus, Trichoderma reesei, Pichia, Penicillium, Streptomyces
CC or Bacillus, and a method for producing recombinant cellulase by
CC culturing these host cells.
XX
XX SQ Sequence 459 AA;
SQ

Query Match 34.7%; Score 41; DB 19; Length 459;
Best Local Similarity 42.9%; Pred. No. 1.8e+02;
Matches 9; Conservative 4; Mismatches 4; Indels 4; Gaps 1;
QY 1 NSPDSCEQQAARNVOHYHKNEL 21
DB 257 nnndnc-----rnvrnmhkqal 273

RESULT 37
AAG07900
ID AAG07900 standard; Protein; 550 AA.
XX
XX AC AAG07900;
XX
XX DT 17-OCT-2000 (first entry)
XX
XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 5227.
XX
XX KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
XX OS Arabidopsis thaliana.
XX
XX PN EP1033405-A2.
XX
XX PD 06-SEP-2000.
XX
XX PF 25-FEB-2000; 2000EP-0301439.
XX

PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126284.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 16-APR-1999; 99US-0128714.
PR 18-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136392.
PR 27-MAY-1999; 99US-0136391.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 17-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140951.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 19-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 27-JUL-1999; 99US-0145951.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 25-AUG-1999; 99US-0150884.
PR 26-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 27-AUG-1999; 99US-0151303.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 13-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.

Wed Dec 19 08:45:36 2001

us-09-462-625-2_copy_55_75_rag

```

PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.
PR 29-OCT-1999; 99US-0162143.

Query Match 34.7%; Score 41; DB 21; Length 550;
Best Local Similarity 41.2%; Fred. No. 2.2e+02;
Matches 7; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

OY 2 SPDSCEQOARNOHYHK 18
Db 61 sskkckefenvykyhk 77

RESULT 38
AAG07899
ID AAG07899 standard; Protein; 575 AA.
XX
AC AAG07899;
XX
XX 17-OCT-2000 (first entry)
XX
XX Arabidopsis thaliana protein fragment SEQ ID NO: 5226.
DE
XX
XX Protein identification; signal transduction pathway; metabolic pathway;
XX hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.
XX
XX Arabidopsis thaliana.
XX
XX EP1033405-A2.
XX
XX
XX 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
XX 05-MAR-1999; 99US-0123380.
XX 09-MAR-1999; 99US-0123348.
XX 23-MAR-1999; 99US-0125788.
XX 25-MAR-1999; 99US-0126264.
XX 29-MAR-1999; 99US-0126785.
XX 01-APR-1999; 99US-0127462.

PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 03-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 17-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
```

PR 19-JUL-1999;	99US-0144332.
PR 19-JUL-1999;	99US-0144333.
PR 19-JUL-1999;	99US-0144334.
PR 19-JUL-1999;	99US-0144335.
PR 20-JUL-1999;	99US-0144352.
PR 20-JUL-1999;	99US-0144632.
PR 20-JUL-1999;	99US-0144884.
PR 21-JUL-1999;	99US-0144814.
PR 21-JUL-1999;	99US-0145086.
PR 21-JUL-1999;	99US-0145088.
PR 22-JUL-1999;	99US-0145085.
PR 22-JUL-1999;	99US-0145087.
PR 22-JUL-1999;	99US-0145089.
PR 22-JUL-1999;	99US-0145192.
PR 23-JUL-1999;	99US-0145145.
PR 23-JUL-1999;	99US-0145218.
PR 23-JUL-1999;	99US-0145224.
PR 26-JUL-1999;	99US-0145276.
PR 27-JUL-1999;	99US-0145913.
PR 27-JUL-1999;	99US-0145918.
PR 27-JUL-1999;	99US-0145919.
PR 28-JUL-1999;	99US-0145951.
PR 02-AUG-1999;	99US-0146386.
PR 02-AUG-1999;	99US-0146388.
PR 02-AUG-1999;	99US-0146389.
PR 03-AUG-1999;	99US-0147038.
PR 04-AUG-1999;	99US-0147204.
PR 04-AUG-1999;	99US-0147302.
PR 05-AUG-1999;	99US-0147192.
PR 05-AUG-1999;	99US-0147260.
PR 06-AUG-1999;	99US-0147303.
PR 09-AUG-1999;	99US-0147416.
PR 09-AUG-1999;	99US-0147493.
PR 09-AUG-1999;	99US-0147935.
PR 10-AUG-1999;	99US-0148171.
PR 11-AUG-1999;	99US-0148319.
PR 12-AUG-1999;	99US-0148341.
PR 13-AUG-1999;	99US-0148565.
PR 13-AUG-1999;	99US-0148684.
PR 16-AUG-1999;	99US-0149368.
PR 17-AUG-1999;	99US-0149175.
PR 18-AUG-1999;	99US-0149426.
PR 20-AUG-1999;	99US-0149722.
PR 20-AUG-1999;	99US-0149723.
PR 20-AUG-1999;	99US-0149929.
PR 23-AUG-1999;	99US-0149902.
PR 23-AUG-1999;	99US-0149930.
PR 25-AUG-1999;	99US-0150566.
PR 26-AUG-1999;	99US-0150884.
PR 27-AUG-1999;	99US-0151065.
PR 27-AUG-1999;	99US-0151066.
PR 27-AUG-1999;	99US-0151080.
PR 30-AUG-1999;	99US-0151303.
PR 31-AUG-1999;	99US-0151438.
PR 01-SEP-1999;	99US-0151930.
PR 07-SEP-1999;	99US-0152363.
PR 10-SEP-1999;	99US-0153070.
PR 13-SEP-1999;	99US-0153758.
PR 13-SEP-1999;	99US-0154018.
PR 16-SEP-1999;	99US-0154039.
PR 20-SEP-1999;	99US-0154779.
PR 22-SEP-1999;	99US-0155139.
PR 23-SEP-1999;	99US-0155486.
PR 24-SEP-1999;	99US-0155659.
PR 28-SEP-1999;	99US-0156458.
PR 29-SEP-1999;	99US-0156596.
PR 04-OCT-1999;	99US-0157117.
PR 05-OCT-1999;	99US-0157753.
PR 06-OCT-1999;	99US-0157865.
PR 07-OCT-1999;	99US-0158029.
PR 08-OCT-1999;	99US-0158232.
PR 12-OCT-1999;	99US-0158369.
PR 13-OCT-1999;	99US-0159293.
PR 13-OCT-1999;	99US-0159294.
PR 13-OCT-1999;	99US-0159295.
PR 14-OCT-1999;	99US-0159329.
PR 14-OCT-1999;	99US-0159330.
PR 14-OCT-1999;	99US-0159331.
PR 14-OCT-1999;	99US-0159637.
PR 14-OCT-1999;	99US-0159638.
PR 18-OCT-1999;	99US-0159584.
PR 21-OCT-1999;	99US-0160741.
PR 21-OCT-1999;	99US-0160767.
PR 21-OCT-1999;	99US-0160768.
PR 21-OCT-1999;	99US-0160770.
PR 21-OCT-1999;	99US-0160814.
PR 21-OCT-1999;	99US-0160815.
PR 22-OCT-1999;	99US-0160980.
PR 22-OCT-1999;	99US-0160981.
PR 22-OCT-1999;	99US-0160989.
PR 25-OCT-1999;	99US-0161404.
PR 25-OCT-1999;	99US-0161405.
PR 25-OCT-1999;	99US-0161406.
PR 26-OCT-1999;	99US-0161359.
PR 26-OCT-1999;	99US-0161360.
PR 26-OCT-1999;	99US-0161361.
PR 28-OCT-1999;	99US-0161920.
PR 28-OCT-1999;	99US-0161992.
PR 28-OCT-1999;	99US-0161993.
PR 29-OCT-1999;	99US-0162142.
Query Match 34.7%; Score 41; DB 21; Length 575;	
Best Local Similarity 41.2%; Pred. No. 2.3e+02;	
Matches 7; Conservative 3; Mismatches 7; Indels 0; Gaps 0;	
QY 2 SPDSCEQARNVQHYHK 18	
Db 86 sskckekfenykyhk 102	
RESULT 39	
AAG07898	
ID AAG07898 standard; Protein; 597 AA.	
XX AC AAG07898;	
XX DT 17-OCT-2000 (first entry)	
XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 5225.	
XX KW Protein identification; signal transduction pathway; metabolic pathway;	
XX KW Hybridisation assay; genetic mapping; gene expression control; promoter;	
XX KW termination sequence.	
XX OS Arabidopsis thaliana.	
XX PN EP1033405-A2.	
XX PD 06-SEP-2000.	
XX PF 25-FEB-2000; 2000EP-0301439.	
XX XX 25-FEB-1999; 99US-0121825.	
PR 03-MAR-1999; 99US-0123180.	
PR 09-MAR-1999; 99US-0123548.	
PR 23-MAR-1999; 99US-0125788.	
PR 25-MAR-1999; 99US-0126264.	
PR 26-MAR-1999; 99US-0126785.	
PR 01-APR-1999; 99US-0127462.	
PR 06-APR-1999; 99US-0128234.	
PR 08-APR-1999; 99US-0128714.	
PR 16-APR-1999; 99US-0129845.	
PR 19-APR-1999; 99US-0130077.	
PR 21-APR-1999; 99US-0130449.	
PR 23-APR-1999; 99US-0130510.	
PR 23-APR-1999; 99US-0130891.	

Wed Dec 19 08:45:36 2001

us-09-462-625-2_copy_55_75.rag

PR	28-APR-1999;	99US-0131449.
PR	30-APR-1999;	99US-0132048.
PR	30-APR-1999;	99US-0132407.
PR	04-MAY-1999;	99US-0132484.
PR	05-MAY-1999;	99US-0132485.
PR	06-MAY-1999;	99US-0132486.
PR	06-MAY-1999;	99US-0132487.
PR	07-MAY-1999;	99US-0132863.
PR	11-MAY-1999;	99US-0134256.
PR	14-MAY-1999;	99US-0134218.
PR	14-MAY-1999;	99US-0134219.
PR	14-MAY-1999;	99US-0134221.
PR	14-MAY-1999;	99US-0134370.
PR	18-MAY-1999;	99US-0134768.
PR	19-MAY-1999;	99US-0134941.
PR	20-MAY-1999;	99US-0135124.
PR	21-MAY-1999;	99US-0135353.
PR	24-MAY-1999;	99US-0135629.
PR	25-MAY-1999;	99US-0136021.
PR	27-MAY-1999;	99US-0136392.
PR	28-MAY-1999;	99US-0136782.
PR	01-JUN-1999;	99US-0137222.
PR	03-JUN-1999;	99US-0137528.
PR	04-JUN-1999;	99US-0137502.
PR	07-JUN-1999;	99US-0137724.
PR	08-JUN-1999;	99US-0138094.
PR	10-JUN-1999;	99US-0138540.
PR	10-JUN-1999;	99US-0138847.
PR	14-JUN-1999;	99US-0139119.
PR	16-JUN-1999;	99US-0139452.
PR	16-JUN-1999;	99US-0139453.
PR	17-JUN-1999;	99US-0139492.
PR	18-JUN-1999;	99US-0139454.
PR	18-JUN-1999;	99US-0139455.
PR	18-JUN-1999;	99US-0139456.
PR	18-JUN-1999;	99US-0139457.
PR	18-JUN-1999;	99US-0139458.
PR	18-JUN-1999;	99US-0139459.
PR	18-JUN-1999;	99US-0139460.
PR	18-JUN-1999;	99US-0139461.
PR	18-JUN-1999;	99US-0139462.
PR	18-JUN-1999;	99US-0139463.
PR	18-JUN-1999;	99US-0139750.
PR	18-JUN-1999;	99US-0139763.
PR	21-JUN-1999;	99US-0139817.
PR	21-JUN-1999;	99US-0139899.
PR	23-JUN-1999;	99US-0140353.
PR	23-JUN-1999;	99US-0140354.
PR	23-JUN-1999;	99US-0140695.
PR	24-JUN-1999;	99US-0140823.
PR	29-JUN-1999;	99US-0140991.
PR	30-JUN-1999;	99US-0141287.
PR	01-JUL-1999;	99US-0141842.
PR	01-JUL-1999;	99US-0142154.
PR	02-JUL-1999;	99US-0142055.
PR	02-JUL-1999;	99US-0142390.
PR	06-JUL-1999;	99US-0142803.
PR	08-JUL-1999;	99US-0142920.
PR	09-JUL-1999;	99US-0142977.
PR	12-JUL-1999;	99US-0143542.
PR	13-JUL-1999;	99US-0143624.
PR	14-JUL-1999;	99US-0144005.
PR	15-JUL-1999;	99US-0144085.
PR	16-JUL-1999;	99US-0144086.
PR	16-JUL-1999;	99US-0144325.
PR	19-JUL-1999;	99US-0144331.
PR	19-JUL-1999;	99US-0144332.
PR	19-JUL-1999;	99US-0144333.
PR	19-JUL-1999;	99US-0144334.
PR	19-JUL-1999;	99US-0144335.
PR	20-JUL-1999;	99US-0144352.
PR	20-JUL-1999;	99US-0144632.
PR	20-JUL-1999;	99US-0144884.
PR	28-APR-1999;	99US-0144814.
PR	21-JUL-1999;	99US-0145086.
PR	21-JUL-1999;	99US-0145088.
PR	22-JUL-1999;	99US-0145085.
PR	22-JUL-1999;	99US-0145087.
PR	22-JUL-1999;	99US-0145089.
PR	22-JUL-1999;	99US-0145192.
PR	22-JUL-1999;	99US-0145145.
PR	23-JUL-1999;	99US-0145218.
PR	23-JUL-1999;	99US-0145224.
PR	26-JUL-1999;	99US-0145276.
PR	27-JUL-1999;	99US-0145913.
PR	27-JUL-1999;	99US-0145918.
PR	27-JUL-1999;	99US-0145919.
PR	27-JUL-1999;	99US-0145951.
PR	28-JUL-1999;	99US-0145951.
PR	02-AUG-1999;	99US-0146386.
PR	02-AUG-1999;	99US-0146388.
PR	02-AUG-1999;	99US-0146389.
PR	03-AUG-1999;	99US-0147038.
PR	04-AUG-1999;	99US-0147204.
PR	04-AUG-1999;	99US-0147302.
PR	05-AUG-1999;	99US-0147192.
PR	06-AUG-1999;	99US-0147260.
PR	06-AUG-1999;	99US-0147303.
PR	06-AUG-1999;	99US-0147416.
PR	03-AUG-1999;	99US-0147493.
PR	03-AUG-1999;	99US-0147935.
PR	10-AUG-1999;	99US-0148171.
PR	11-AUG-1999;	99US-0148319.
PR	12-AUG-1999;	99US-0148341.
PR	13-AUG-1999;	99US-0148565.
PR	13-AUG-1999;	99US-0148684.
PR	16-AUG-1999;	99US-0149368.
PR	17-AUG-1999;	99US-0149175.
PR	18-AUG-1999;	99US-0149426.
PR	20-AUG-1999;	99US-0149722.
PR	20-AUG-1999;	99US-0149723.
PR	20-AUG-1999;	99US-0149929.
PR	23-AUG-1999;	99US-0149902.
PR	23-AUG-1999;	99US-0149930.
PR	25-AUG-1999;	99US-0150566.
PR	26-AUG-1999;	99US-0150884.
PR	27-AUG-1999;	99US-0151065.
PR	27-AUG-1999;	99US-0151066.
PR	30-AUG-1999;	99US-0151080.
PR	30-AUG-1999;	99US-0151303.
PR	31-AUG-1999;	99US-0151438.
PR	01-SEP-1999;	99US-0151930.
PR	07-SEP-1999;	99US-0152363.
PR	10-SEP-1999;	99US-0153070.
PR	13-SEP-1999;	99US-0153758.
PR	15-SEP-1999;	99US-0154018.
PR	16-SEP-1999;	99US-0154039.
PR	20-SEP-1999;	99US-0154779.
PR	23-SEP-1999;	99US-0155139.
PR	23-SEP-1999;	99US-0155486.
PR	28-SEP-1999;	99US-0155659.
PR	28-SEP-1999;	99US-0156458.
PR	29-SEP-1999;	99US-0156596.
PR	04-OCT-1999;	99US-0157117.
PR	05-OCT-1999;	99US-0157533.
PR	06-OCT-1999;	99US-0157865.
PR	07-OCT-1999;	99US-0158029.
PR	08-OCT-1999;	99US-0158232.
PR	12-OCT-1999;	99US-0158369.
PR	13-OCT-1999;	99US-0159293.
PR	13-OCT-1999;	99US-0159294.
PR	13-OCT-1999;	99US-0159295.
PR	14-OCT-1999;	99US-0159329.
PR	14-OCT-1999;	99US-0159330.
PR	14-OCT-1999;	99US-0159331.
PR	14-OCT-1999;	99US-0159637.
PR	14-OCT-1999;	99US-0159638.

PR 18-OCT-1999; 99US-0159584.
 PR 21-OCT-1999; 99US-0160741.
 PR 21-OCT-1999; 99US-0160767.
 PR 21-OCT-1999; 99US-0160768.
 PR 21-OCT-1999; 99US-0160770.
 PR 21-OCT-1999; 99US-0160814.
 PR 21-OCT-1999; 99US-0160815.
 PR 22-OCT-1999; 99US-0160980.
 PR 22-OCT-1999; 99US-0160981.
 PR 22-OCT-1999; 99US-0160989.
 PR 22-OCT-1999; 99US-0161404.
 PR 25-OCT-1999; 99US-0161405.
 PR 25-OCT-1999; 99US-0161406.
 PR 26-OCT-1999; 99US-0161359.
 PR 26-OCT-1999; 99US-0161360.
 PR 26-OCT-1999; 99US-0161361.
 PR 28-OCT-1999; 99US-0161920.
 PR 28-OCT-1999; 99US-0161992.
 PR 28-OCT-1999; 99US-0161993.
 PR 29-OCT-1999; 99US-0162142.

Query Match 34.7%; Score 41; DB 21; Length 597;
 Best Local Similarity 41.2%; Pred. No. 2.4e+02;
 Matches 7; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

OY 2 SPDSCEQOARNVQHYHK 18
 | : : : | : : :
 Db 108 sskckekfenvkyhk 124

RESULT 40
 AAY22176
 ID AAY22176 standard; Protein; 716 AA.
 XX
 AC AAY22176;
 XX
 DT 09-SEP-1999 (first entry)
 XX
 DE Drosophila Acp36DE protein.
 XX
 KW Accessory gland protein; Acp; toxin; insecticide; Drosophila; mating;
 KW caterpillar; development inhibitor; insect pest; plant protection.
 XX
 OS Drosophila melanogaster.
 XX
 PN WO9932149-A1.
 XX
 PD 01-JUL-1999.
 XX
 PF 23-DEC-1998; 98WO-US27603.
 XX
 PR 23-DEC-1997; 97US-0071315.
 XX
 PA (CORR) CORNELL RES FOUND INC.
 XX
 PI Lung O, Tram K, Wolfner MF;
 XX
 DR WPI; 1999-418871/35.
 DR N-PSDB; AAX84367.
 XX
 PT Nucleic acid encoding accessory gland proteins of Drosophila
 XX
 PS Claim 38; Page 24-25; 89pp; English.
 XX
 CC This sequence is a Drosophila melanogaster accessory gland protein
 CC (Acp) of the invention. A particular Acp, designated Acp62F, is toxic to
 CC insects, particularly to Drosophila and caterpillars, and it (or vectors
 CC that express it) can be used to kill or inhibit development of insect
 CC pests, for plant protection. More generally detection of Acp's in a
 CC female fruit fly is indicative of recent mating.
 XX
 SQ Sequence 716 AA;

Query Match 34.7%; Score 41; DB 20; Length 716;
 Best Local Similarity 37.5%; Pred. No. 3e+02;
 Matches 9; Conservative 5; Mismatches 6; Indels 4; Gaps 1;

OY 2 SPDSCEQO----ARNVQHYHKNEL 21
 | | | | : : : : | : : :
 Db 303 spgglegqllhlqullmfqgnql 326

RESULT 41
 AAY54095
 ID AAY54095 standard; Protein; 327 AA.
 XX
 AC AAY54095;
 XX
 DT 27-MAR-2000 (first entry)
 XX
 DE Enzyme EPSJ involved in exopolysaccharide biosynthesis.
 XX
 KW Exopolysaccharide; EPS; ESP enzyme; EPSA; EPSB; EPSD; EPSE; EPSF;
 KW EPSG; EPSI; EPSJ; EPSK; Lactobacillus delbrueckii bulgaricus;
 KW activated D-galactose pyranose; saccharide; beta-glycosyltransferase;
 KW transcription attenuator; glucosyl-phospho-transferase;
 KW alpha-glycosyltransferase; glucosyltransferase; EPSL; EPSM;
 KW alpha-glycosyltransferase; EPS polymerase; glycosyltransferase; EPSN;
 KW phosphofuranose; transporter; food; fermented milk product; yoghurt;
 KW cheese; flavour stability; organoleptic property.
 XX
 OS Lactobacillus delbrueckii bulgaricus.
 XX
 PN WO9962316-A2.
 XX
 PD 09-DEC-1999.
 XX
 PF 22-APR-1999; 99WO-EP02841.
 XX
 PR 22-APR-1998; 98EP-0201310.
 PR 22-APR-1998; 98EP-0201311.
 PR 22-APR-1998; 98EP-0201312.
 XX
 PA (NEST) SOC PROD NESTLE SA.
 XX
 PI Stinglele F, Germond JE, Lamothe G;
 XX
 DR WPI; 2000-097267/08.
 DR N-PSDB; AAZ45260.
 XX
 PT New recombinant enzymes for synthesis of exopolysaccharides,
 PT particularly in lactic acid bacteria, for improving properties of
 PT fermented milk products -
 XX
 PS Example 18; Page 156-157; 162pp; French.
 XX
 CC AAY54096-99 represent enzymes involved in the biosynthesis of
 CC exopolysaccharides (EPS). These enzymes are designated EPSA-EPSN, and
 CC are encoded by open reading frames epsA-epsN. The enzymes are isolated
 CC from Lactobacillus delbrueckii bulgaricus. The proteins are used
 CC in a method for the synthesis of EPS, which includes at least one step
 CC of forming a bond (alpha or beta-isomer) between C-1 (carrying the
 CC reducing aldehyde function, of an activated D-galactose pyranose), and
 CC a phosphate on a lipophilic or proteinaceous primer. Stepwise synthesis
 CC of EPS occurs with, in each step, addition of a new sugar unit, through
 CC its hemi-acetyl function, to an alcoholic hydroxyl of a second sugar
 CC unit, present at the end of a chain of sugar residues bonded to the
 CC primer. EPSA is attenuator of transcription which influences
 CC regulation; EPSB determines the length of the saccharide chain; EPSK is
 CC involved in the control of the molecular weight and/or the length of the
 CC polysaccharide; EPSD, EPSL and EPSM are involved in synthesis of EPS;
 CC EPSE is a galactosyl- or glucosyl-phospho-transferase which catalyses the
 CC transfer of the first saccharide on the primer; EPSF and EPSG are
 CC alpha-glycosyltransferases; EPSH and EPSJ are
 CC beta-glycosyltransferases; EPSI is a glucosyltransferase; EPSK is

CC responsible for the polymerisation of the repetitive units; and EPSN is
CC responsible for export of the EPS. The EPS enzyme are used to improve
CC properties of foods, particularly fermented milk products such as yoghurt
CC and cheese, e.g. their organoleptic properties and flavour stability.
XX
SQ Sequence 327 AA;

Query Match 34.3%; Score 40.5; DB 21; Length 327;
Best Local Similarity 41.7%; Pred. No. 1.5e+02;
Matches 10; Conservative 2; Mismatches 5; Indels 7; Gaps 1;

OY 2 SPDSC-----EQQARNVOHYHK 18
||||| : : : :
Db 40 spdscpkicdeysqkfnvkvvhk 63

RESULT 42
AAV43797
ID AAY43797 standard; Protein; 327 AA.
XX
AC
XX
XX AAY43797;
DT 11-FEB-2000 (first entry)
DE Amino acid sequence of epsJ of L. delbrueckii bulgaricus Lfi5.
XX
XX eps operon; Lactobacillus delbrueckii bulgaricus Lfi5; enzyme; epsA;
KW epsB; epsC; epsD; epsE; epsF; epsG; epsH; epsI; epsJ; epsK; epsL; epsM;
KW exopolysaccharide biosynthesis; EPS; intersugar bond; antitumour;
KW probiotic; foodstuff; organoleptic quality; flavour;
KW lactic acid bacteria; acidified milk product; yoghurt; cheese.
XX
XX Lactobacillus delbrueckii bulgaricus.
OS
XX
XX WO9954475-A2.
XX
XX 28-OCT-1999.
XX
XX 22-APR-1999; 99WO-EP03011.
XX
XX 22-APR-1998; 98EP-0201310.
XX
XX 22-APR-1998; 98EP-0201311.
XX
XX 22-APR-1998; 98EP-0201312.
XX
XX (NEST) SOC PROD NESTLE SA.
XX
XX Stingle F, Germond JE, Lamothe G;
PI
XX
XX WPI; 2000-013255/01.
XX
XX N-PSDB; AAZ30357, AAY43796, AAY43797, AAY43798, AAY43799, AAY43800,
XX
XX AAY43801, AAY43842.
XX
XX New recombinant enzymes for biosynthesis of exopolysaccharides having
XX
XX e.g. antitumor or probiotic properties or useful in fermented milk
XX
XX products -
XX
XX Example 18; Page 157-158; 163pp; French.

XX
XX AAY43798-89, AAY43791-Y437801 and AAY43842 represent the enzymes encoded
XX
XX by the eps operon of Lactobacillus delbrueckii bulgaricus Lfi5. The
XX
XX operon contains 14 open reading frames, and encodes enzymes (epsA,
XX
XX epsB, epsC, epsD, epsE, epsF, epsG, epsH, epsI and epsJ, epsK, epsL,
XX
XX and epsM) that are involved in the biosynthesis of exopolysaccharides
XX
XX (EPS). The enzymes catalyse the formation of specific intersugar bonds.
XX
XX The enzymes catalyse a process that includes at least one step of
XX
XX forming a bond (in alpha or beta anomeric form) between C1, carrying
XX
XX the reducing aldehyde group of an activated D-Galp (galactose in
XX
XX pyranose form), and a phosphate group on a lipophilic or proteinaceous
XX
XX primer. The enzymes are used to produce EPS that have antitumor or
XX
XX probiotic properties or are used in foodstuffs to improve organoleptic
XX
XX qualities and flavour. When expressed by lactic acid bacteria, EPS
XX
XX impart a free-flowing character and/or a smooth, creamy texture to

CC acidified milk products (yoghurt or cheese).
XX
SQ Sequence 327 AA;

Query Match 34.3%; Score 40.5; DB 21; Length 327;
Best Local Similarity 41.7%; Pred. No. 1.5e+02;
Matches 10; Conservative 2; Mismatches 5; Indels 7; Gaps 1;

OY 2 SPDSC-----EQQARNVOHYHK 18
||||| : : : :
Db 40 spdscpkicdeysqkfnvkvvhk 63

RESULT 43
AAB38431
ID AAB38431 standard; peptide; 40 AA.
XX
XX
AC AAB38431;
XX
XX 31-JAN-2001 (first entry)
DT
DE Fragment of human secreted protein encoded by gene 10 clone HTEBV72.
XX
XX
XX Immunosuppressive; antiarthritic; antirheumatic; antiproliferative;
KW cytostatic; cardiant; vasotropic; cerebroprotective; neuroprotective;
KW neurotropic; antibacterial; virucide; fungicide; ophthalmological; human;
KW vulnery; gene therapy; infection; secreted protein.
XX
XX Homo sapiens.
OS
XX
XX WO2000061623-A1.
XX
XX 19-OCT-2000.
XX
XX 06-APR-2000; 2000WO-US08979.
XX
XX 09-APR-1999; 99US-0128693.
XX
XX 26-APR-1999; 99US-0130991.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Ruben SM, Ni J, Komatsoulis GA, Rosen CA, Soppet DR, Shi Y;
XX
XX Lafleur DW, Olsen HS, Ebner R, Florence KA, Moore PA, Birse CE;
XX
XX Young PE;
XX
XX WPI; 2000-647418/62.

XX
XX New nucleic acid molecules encoding 62 human secreted proteins for
XX
XX diagnosing, preventing, treating or ameliorating medical conditions and
XX
XX used as food additives or preservatives -
XX
XX Disclosure; Page 32; 716pp; English.

XX
XX Sequences AAB38321-B38396 represent the amino acid sequences of 62
XX
XX human secreted proteins encoded by the genes AAC69512-C69587. The genes
XX
XX and proteins are useful for preventing, ameliorating or treating medical
XX
XX conditions, e.g. by protein or gene therapy. The genes are isolated from
XX
XX a range of human tissues disclosed in the specification. The nucleic
XX
XX acids, proteins, antibodies and (ant)agonists are useful in the
XX
XX diagnosis, treatment and prevention of: (a) autoimmune diseases e.g.
XX
XX rheumatoid arthritis; (b) hyperproliferative disorders e.g. cardiac
XX
XX arrest; (d) cerebrovascular disorders e.g. cerebral ischemia; (e)
XX
XX angiogenesis; (f) nervous system disorders e.g. Alzheimer's disease; (g)
XX
XX infections caused by bacteria, viruses and fungi; and (h) ocular
XX
XX disorders e.g. corneal infection. The polypeptides can also be used to
XX
XX aid wound healing and epithelial cell proliferation, to prevent skin
XX
XX aging due to sunburn, to maintain organs before transplantation, for
XX
XX supporting cell culture of primary tissues, to regenerate tissues and in
XX
XX chemotaxis.

XX
XX Sequence 40 AA;

Query Match 33.9%; Score 40; DB 21; Length 40;
Best Local Similarity 31.2%; Pred. No. 16;
Matches 5; Conservative 7; Mismatches 4; Indels 0; Gaps 0;

QY 2 SPDSCEQQARNVQHYH 17
:|: |: |::|:|
Db 24 tpevieksvrdlehw 39

RESULT 44
AAB38428
ID AAB38428 standard; peptide; 171 AA.
XX AC
XX AAB38428;
XX 31-JAN-2001 (first entry)
XX Fragment of human secreted protein encoded by gene 10 clone HTERV72.
XX Immunosuppressive; antiarthritic; antirheumatic; antiproliferative;
KW cytostatic; cardiant; vasotropic; cerebroprotective; neuroprotective;
KW neotropic; antibacterial; virucide; fungicide; opthalmological; human;
KW vulnary; gene therapy; infection; secreted protein.
XX
XX Homo sapiens.
XX WO200061623-A1.
XX PN
XX 19-OCT-2000.
XX PD
XX PF 06-APR-2000; 2000WO-US08979.
XX PR 09-APR-1999; 99US-0128693.
XX PR 26-APR-1999; 99US-0130991.
XX PA (HUMA-) HUMAN GENOME SCI INC.
XX Ruben SM, NI J, Komatsoulis GA, Rosen CA, Soppet DR, Shi Y;
PI Lafleur DW, Olsen HS, Ebner R, Florence KA, Moore PA, Birse CE;
PI Young PE;
XX WPI; 2000-647418/62.
XX New nucleic acid molecules encoding 62 human secreted proteins for
PT diagnosing, preventing, treating or ameliorating medical conditions and
PT used as food additives or preservatives -
PS Disclosure; Page 32; 716pp; English.
XX Sequences AAB38321-B38396 represent the amino acid sequences of 62
CC human secreted proteins encoded by the genes AAC69512-C69587. The genes
CC and proteins are useful for preventing, ameliorating or treating medical
CC conditions, e.g. by protein or gene therapy. The genes are isolated from
CC a range of human tissues disclosed in the specification. The nucleic
CC acids, proteins, antibodies and (ant)agonists are useful in the
CC diagnosis, treatment and prevention of: (a) autoimmune diseases e.g.
CC rheumatoid arthritis; (b) hyperproliferative disorders e.g. neoplasms
CC of the breast or liver; (c) cardiovascular disorders e.g. cardiac
CC arrest; (d) cerebrovascular disorders e.g. cerebral ischemia; (e)
CC angiogenesis; (f) nervous system disorders e.g. Alzheimer's disease; (g)
CC infections caused by bacteria, viruses and fungi; and (h) ocular
CC disorders e.g. corneal infection. The polypeptides can also be used to
CC aid wound healing and epithelial cell proliferation, to prevent skin
CC aging due to sunburn, to maintain organs before transplantation, for
CC supporting cell culture of primary tissues, to regenerate tissues and in
CC chemotaxis.
XX Sequence 171 AA;

Query Match 33.9%; Score 40; DB 21; Length 171;

Best Local Similarity 31.2%; Pred. No. 83;
Matches 5; Conservative 7; Mismatches 4; Indels 0; Gaps 0;

QY 2 SPDSCEQQARNVQHYH 17
:|: |: |::|:|
Db 117 tpevieksvrdlehw 132

RESULT 45
AAR43385
ID AAR43385 standard; Protein; 231 AA.
XX AC
XX AAR43385;
XX 19-JUN-1994 (first entry)
XX Product of homeotic gene "green petal".
XX Plant; organ morphogenesis; control; petunia; petals.
XX Petunia.
XX WO9321322-A.
XX PD 28-OCT-1993.
XX PF 13-APR-1993; 93WO-US03508.
XX PR 13-APR-1992; 92US-0867580.
XX PR 06-JUL-1992; 92US-0909589.
XX PA (UYRQ) UNIV ROCKEFELLER.
XX Chua N, Halfter U, Kush A, Van DER KROL AR;
PI WPI; 1993-351732/44.
DR N-PSDB; AAQ51189.
XX PT Plant organ morphogenesis control and determ. - by regulating
PT the expression of homeotic genes which determine the identity of
PT the organ
XX Disclosure; Fig 2; 74pp; English.
XX The homeotic gene green petal from petunia has been cloned and
CC characterised previously. The gene was used in a new method for
CC controlling the morphogenesis of plant organs comprising regulating
CC the expression of the gene using ectopic expression. Such a method
CC can be used to determine and control plant organ morphogenesis, such
CC as modifying petals without altering the reproductive portions of
CC the flower.
CC See also AAR43386-7.
XX Sequence 231 AA;

Query Match 33.9%; Score 40; DB 14; Length 231;
Best Local Similarity 33.3%; Pred. No. 1.2e+02;
Matches 7; Conservative 7; Mismatches 7; Indels 0; Gaps 0;

QY 1 NSPDSCEQQARNVQHYHKNEL 21
:|: |: |::|:|
Db 145 nqietfkkkvrnveelhrnl 165

RESULT 46
AAG59547
ID AAG59547 standard; Protein; 246 AA.
XX AC
XX AAG59547;
XX 18-OCT-2000 (first entry)
XX

DE Arabidopsis thaliana protein fragment SEQ ID NO: 77032.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
OS Arabidopsis thaliana.
XX
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 26-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 16-AUG-1999; 99US-0149175.
PR 17-AUG-1999; 99US-0149426.
PR 18-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.

PR	30-AUG-1999;	99US-0151303.	XX	EP1033405-A2.	
PR	31-AUG-1999;	99US-0151438.	PN		
PR	01-SEP-1999;	99US-0151930.	XX		
PR	07-SEP-1999;	99US-0152363.	PD	06-SEP-2000.	
PR	10-SEP-1999;	99US-0153070.	XX		
PR	13-SEP-1999;	99US-0153758.	XX	25-FEB-2000;	2000EP-0301439.
PR	15-SEP-1999;	99US-0154018.	XX		
PR	16-SEP-1999;	99US-0154039.	PR	25-FEB-1999;	99US-0121825.
PR	20-SEP-1999;	99US-0154779.	PR	05-MAR-1999;	99US-0123180.
PR	22-SEP-1999;	99US-0155139.	PR	09-MAR-1999;	99US-0123548.
PR	23-SEP-1999;	99US-0155486.	PR	23-MAR-1999;	99US-0125788.
PR	24-SEP-1999;	99US-0155659.	PR	25-MAR-1999;	99US-0126264.
PR	28-SEP-1999;	99US-0156458.	PR	29-MAR-1999;	99US-0126785.
PR	29-SEP-1999;	99US-0156596.	PR	01-APR-1999;	99US-0127462.
PR	04-OCT-1999;	99US-0157117.	PR	06-APR-1999;	99US-0128234.
PR	05-OCT-1999;	99US-0157753.	PR	08-APR-1999;	99US-0128714.
PR	06-OCT-1999;	99US-0157865.	PR	16-APR-1999;	99US-0129845.
PR	07-OCT-1999;	99US-0158029.	PR	19-APR-1999;	99US-0130077.
PR	08-OCT-1999;	99US-0158232.	PR	21-APR-1999;	99US-0130449.
PR	12-OCT-1999;	99US-0158369.	PR	23-APR-1999;	99US-0130510.
PR	13-OCT-1999;	99US-0159293.	PR	23-APR-1999;	99US-0130891.
PR	13-OCT-1999;	99US-0159294.	PR	28-APR-1999;	99US-0131449.
PR	13-OCT-1999;	99US-0159295.	PR	30-APR-1999;	99US-0132048.
PR	14-OCT-1999;	99US-0159330.	PR	04-MAY-1999;	99US-0132407.
PR	14-OCT-1999;	99US-0159331.	PR	05-MAY-1999;	99US-0132484.
PR	14-OCT-1999;	99US-0159637.	PR	06-MAY-1999;	99US-0132486.
PR	18-OCT-1999;	99US-0159584.	PR	07-MAY-1999;	99US-0132487.
PR	21-OCT-1999;	99US-0160741.	PR	11-MAY-1999;	99US-0132863.
PR	21-OCT-1999;	99US-0160767.	PR	14-MAY-1999;	99US-0134256.
PR	21-OCT-1999;	99US-0160768.	PR	14-MAY-1999;	99US-0134219.
PR	21-OCT-1999;	99US-0160814.	PR	14-MAY-1999;	99US-0134221.
PR	21-OCT-1999;	99US-0160815.	PR	18-MAY-1999;	99US-0134370.
PR	22-OCT-1999;	99US-0160980.	PR	19-MAY-1999;	99US-0134768.
PR	22-OCT-1999;	99US-0160981.	PR	20-MAY-1999;	99US-0134941.
PR	22-OCT-1999;	99US-0160989.	PR	21-MAY-1999;	99US-0135124.
PR	25-OCT-1999;	99US-0161404.	PR	24-MAY-1999;	99US-0135353.
PR	25-OCT-1999;	99US-0161405.	PR	25-MAY-1999;	99US-0135629.
PR	26-OCT-1999;	99US-0161359.	PR	27-MAY-1999;	99US-0136021.
PR	26-OCT-1999;	99US-0161360.	PR	28-MAY-1999;	99US-0136392.
PR	26-OCT-1999;	99US-0161361.	PR	01-JUN-1999;	99US-0136782.
PR	28-OCT-1999;	99US-0161920.	PR	03-JUN-1999;	99US-0137222.
PR	28-OCT-1999;	99US-0161921.	PR	04-JUN-1999;	99US-0137528.
PR	28-OCT-1999;	99US-0161993.	PR	07-JUN-1999;	99US-0137502.
PR	29-OCT-1999;	99US-0162142.	PR	08-JUN-1999;	99US-0137724.
PR			PR	10-JUN-1999;	99US-0138094.
			PR	10-JUN-1999;	99US-0138540.
			PR	10-JUN-1999;	99US-0138847.
			PR	14-JUN-1999;	99US-0139119.
			PR	16-JUN-1999;	99US-0139452.
			PR	16-JUN-1999;	99US-0139453.
			PR	17-JUN-1999;	99US-0139492.
			PR	18-JUN-1999;	99US-0139454.
			PR	18-JUN-1999;	99US-0139455.
			PR	18-JUN-1999;	99US-0139456.
			PR	18-JUN-1999;	99US-0139457.
			PR	18-JUN-1999;	99US-0139458.
			PR	18-JUN-1999;	99US-0139459.
			PR	18-JUN-1999;	99US-0139460.
			PR	18-JUN-1999;	99US-0139461.
			PR	18-JUN-1999;	99US-0139462.
			PR	18-JUN-1999;	99US-0139463.
			PR	18-JUN-1999;	99US-0139750.
			PR	18-JUN-1999;	99US-0139763.
			PR	21-JUN-1999;	99US-0139817.
			PR	22-JUN-1999;	99US-0139899.
			PR	23-JUN-1999;	99US-0140353.
			PR	23-JUN-1999;	99US-0140354.
			PR	24-JUN-1999;	99US-0140695.
			PR	28-JUN-1999;	99US-0140823.
			PR	29-JUN-1999;	99US-0140991.
			PR	30-JUN-1999;	99US-0141287.

Query Match 33.9%; Score 40; DB 21; Length 246;
Best Local Similarity 41.2%; Pred. No. 1.3e+02;
Matches 7; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

QY 4 DSCEQQARNVQVHKNE 20
| :||| | | :|
Db 102 dvlgkarnyhhlpq 118

RESULT 47
AAG59272
ID AAG59272 standard; Protein; 250 AA.
XX
AC AAG59272;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 76651.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.

[illegible]

PR 25-FEB-1999;	99US-0121825.	PR 13-JUL-1999;	99US-0143542.
PR 05-MAR-1999;	99US-0123180.	PR 14-JUL-1999;	99US-0143624.
PR 09-MAR-1999;	99US-0123548.	PR 15-JUL-1999;	99US-0144005.
PR 23-MAR-1999;	99US-0125788.	PR 16-JUL-1999;	99US-0144085.
PR 25-MAR-1999;	99US-0126264.	PR 19-JUL-1999;	99US-0144086.
PR 29-MAR-1999;	99US-0126785.	PR 19-JUL-1999;	99US-0144325.
PR 01-APR-1999;	99US-0127462.	PR 19-JUL-1999;	99US-0144331.
PR 06-APR-1999;	99US-0128234.	PR 19-JUL-1999;	99US-0144332.
PR 08-APR-1999;	99US-0128714.	PR 19-JUL-1999;	99US-0144334.
PR 16-APR-1999;	99US-0129845.	PR 19-JUL-1999;	99US-0144335.
PR 19-APR-1999;	99US-0130077.	PR 20-JUL-1999;	99US-0144352.
PR 21-APR-1999;	99US-0130449.	PR 20-JUL-1999;	99US-0144632.
PR 23-APR-1999;	99US-0130510.	PR 20-JUL-1999;	99US-0144884.
PR 23-APR-1999;	99US-0130891.	PR 21-JUL-1999;	99US-0144814.
PR 28-APR-1999;	99US-0131449.	PR 21-JUL-1999;	99US-0145086.
PR 30-APR-1999;	99US-0132048.	PR 21-JUL-1999;	99US-0145088.
PR 30-APR-1999;	99US-0132407.	PR 22-JUL-1999;	99US-0145085.
PR 04-MAY-1999;	99US-0132484.	PR 22-JUL-1999;	99US-0145087.
PR 05-MAY-1999;	99US-0132485.	PR 22-JUL-1999;	99US-0145089.
PR 06-MAY-1999;	99US-0132486.	PR 22-JUL-1999;	99US-0145192.
PR 07-MAY-1999;	99US-0132487.	PR 23-JUL-1999;	99US-0145145.
PR 11-MAY-1999;	99US-0132863.	PR 23-JUL-1999;	99US-0145218.
PR 14-MAY-1999;	99US-0134256.	PR 23-JUL-1999;	99US-0145224.
PR 14-MAY-1999;	99US-0134218.	PR 26-JUL-1999;	99US-0145276.
PR 14-MAY-1999;	99US-0134219.	PR 27-JUL-1999;	99US-0145913.
PR 14-MAY-1999;	99US-0134221.	PR 27-JUL-1999;	99US-0145918.
PR 14-MAY-1999;	99US-0134370.	PR 27-JUL-1999;	99US-0145919.
PR 18-MAY-1999;	99US-0134376.	PR 28-JUL-1999;	99US-0145951.
PR 19-MAY-1999;	99US-0134941.	PR 02-AUG-1999;	99US-0146386.
PR 20-MAY-1999;	99US-0135124.	PR 02-AUG-1999;	99US-0146388.
PR 21-MAY-1999;	99US-0135353.	PR 02-AUG-1999;	99US-0146389.
PR 24-MAY-1999;	99US-0135629.	PR 03-AUG-1999;	99US-0147038.
PR 25-MAY-1999;	99US-0136021.	PR 04-AUG-1999;	99US-0147204.
PR 27-MAY-1999;	99US-0136392.	PR 04-AUG-1999;	99US-0147302.
PR 28-MAY-1999;	99US-0136782.	PR 05-AUG-1999;	99US-0147192.
PR 01-JUN-1999;	99US-0137222.	PR 05-AUG-1999;	99US-0147260.
PR 03-JUN-1999;	99US-0137528.	PR 06-AUG-1999;	99US-0147303.
PR 04-JUN-1999;	99US-0137502.	PR 06-AUG-1999;	99US-0147416.
PR 07-JUN-1999;	99US-0137724.	PR 09-AUG-1999;	99US-0147493.
PR 08-JUN-1999;	99US-0138094.	PR 09-AUG-1999;	99US-0147935.
PR 10-JUN-1999;	99US-0138540.	PR 10-AUG-1999;	99US-0148171.
PR 10-JUN-1999;	99US-0138847.	PR 11-AUG-1999;	99US-0148319.
PR 14-JUN-1999;	99US-0139119.	PR 12-AUG-1999;	99US-0148341.
PR 16-JUN-1999;	99US-0139453.	PR 13-AUG-1999;	99US-0148565.
PR 17-JUN-1999;	99US-0139492.	PR 16-AUG-1999;	99US-0148684.
PR 18-JUN-1999;	99US-0139454.	PR 17-AUG-1999;	99US-0149175.
PR 18-JUN-1999;	99US-0139455.	PR 18-AUG-1999;	99US-0149426.
PR 18-JUN-1999;	99US-0139456.	PR 20-AUG-1999;	99US-0149722.
PR 18-JUN-1999;	99US-0139457.	PR 20-AUG-1999;	99US-0149723.
PR 18-JUN-1999;	99US-0139458.	PR 20-AUG-1999;	99US-0149929.
PR 18-JUN-1999;	99US-0139459.	PR 23-AUG-1999;	99US-0149902.
PR 18-JUN-1999;	99US-0139460.	PR 25-AUG-1999;	99US-0149930.
PR 18-JUN-1999;	99US-0139461.	PR 26-AUG-1999;	99US-0150566.
PR 18-JUN-1999;	99US-0139462.	PR 27-AUG-1999;	99US-0150884.
PR 18-JUN-1999;	99US-0139463.	PR 27-AUG-1999;	99US-0151065.
PR 18-JUN-1999;	99US-0139750.	PR 27-AUG-1999;	99US-0151066.
PR 18-JUN-1999;	99US-0139763.	PR 30-AUG-1999;	99US-0151080.
PR 21-JUN-1999;	99US-0139817.	PR 31-AUG-1999;	99US-0151303.
PR 22-JUN-1999;	99US-0139899.	PR 01-SEP-1999;	99US-0151438.
PR 23-JUN-1999;	99US-0140353.	PR 07-SEP-1999;	99US-0151930.
PR 23-JUN-1999;	99US-0140354.	PR 10-SEP-1999;	99US-0152363.
PR 24-JUN-1999;	99US-0140695.	PR 13-SEP-1999;	99US-0153070.
PR 28-JUN-1999;	99US-0140823.	PR 15-SEP-1999;	99US-0153758.
PR 28-JUN-1999;	99US-0140991.	PR 15-SEP-1999;	99US-0154018.
PR 30-JUN-1999;	99US-0141287.	PR 16-SEP-1999;	99US-0154039.
PR 01-JUL-1999;	99US-0141842.	PR 20-SEP-1999;	99US-0154779.
PR 01-JUL-1999;	99US-0142154.	PR 22-SEP-1999;	99US-0155139.
PR 02-JUL-1999;	99US-0142055.	PR 23-SEP-1999;	99US-0155486.
PR 06-JUL-1999;	99US-0142390.	PR 24-SEP-1999;	99US-0155659.
PR 08-JUL-1999;	99US-0142803.	PR 28-SEP-1999;	99US-0156458.
PR 09-JUL-1999;	99US-0142920.	PR 29-SEP-1999;	99US-0156596.
PR 12-JUL-1999;	99US-0142977.		

```
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0158293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 22-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.
PR 29-OCT-1999; 99US-0162142.

Query Match 33.98; Score 40; DB 21; Length 250;
Best Local Similarity 41.28; Pred. No. 1.3e+02;
Matches 7; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

OY 4 DSCQOARNVQHYHNE 20
Db 106 dvlgkarnyhhbq 122
| :||| | :| :
| :||| | :| :

RESULT 49
AAG19303
ID AAG19303 standard; Protein; 262 AA.
XX
AC AAG19303;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 21053.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.

106-APR-1999; 99US-0128234.
108-APR-1999; 99US-0128714.
116-APR-1999; 99US-0129845.
119-APR-1999; 99US-0130077.
121-APR-1999; 99US-0130449.
123-APR-1999; 99US-0130510.
123-APR-1999; 99US-0130891.
123-APR-1999; 99US-0131449.
130-APR-1999; 99US-0132048.
130-APR-1999; 99US-0132407.
130-APR-1999; 99US-0132484.
104-MAY-1999; 99US-0132485.
105-MAY-1999; 99US-0132486.
106-MAY-1999; 99US-0132487.
107-MAY-1999; 99US-0132487.
111-MAY-1999; 99US-0132863.
111-MAY-1999; 99US-0132863.
114-MAY-1999; 99US-0134218.
114-MAY-1999; 99US-0134219.
114-MAY-1999; 99US-0134221.
114-MAY-1999; 99US-0134370.
118-MAY-1999; 99US-0134768.
119-MAY-1999; 99US-0134941.
120-MAY-1999; 99US-0135124.
121-MAY-1999; 99US-0135353.
121-MAY-1999; 99US-0135629.
125-MAY-1999; 99US-0136021.
127-MAY-1999; 99US-0136392.
128-MAY-1999; 99US-0136782.
101-JUN-1999; 99US-0137222.
103-JUN-1999; 99US-0137528.
104-JUN-1999; 99US-0137528.
104-JUN-1999; 99US-0137724.
108-JUN-1999; 99US-0138094.
110-JUN-1999; 99US-0138540.
110-JUN-1999; 99US-0138847.
114-JUN-1999; 99US-0139119.
116-JUN-1999; 99US-0139452.
116-JUN-1999; 99US-0139453.
117-JUN-1999; 99US-0139492.
118-JUN-1999; 99US-0139454.
118-JUN-1999; 99US-0139455.
118-JUN-1999; 99US-0139456.
118-JUN-1999; 99US-0139457.
118-JUN-1999; 99US-0139458.
118-JUN-1999; 99US-0139459.
118-JUN-1999; 99US-0139460.
118-JUN-1999; 99US-0139461.
118-JUN-1999; 99US-0139462.
118-JUN-1999; 99US-0139463.
118-JUN-1999; 99US-0139750.
118-JUN-1999; 99US-0139763.
121-JUN-1999; 99US-0139817.
122-JUN-1999; 99US-0139899.
123-JUN-1999; 99US-0140353.
123-JUN-1999; 99US-0140354.
124-JUN-1999; 99US-0140695.
128-JUN-1999; 99US-0140823.
129-JUN-1999; 99US-0140991.
130-JUN-1999; 99US-0141287.
101-JUL-1999; 99US-0141842.
101-JUL-1999; 99US-0142154.
102-JUL-1999; 99US-0142053.
106-JUL-1999; 99US-0142390.
108-JUL-1999; 99US-0142803.
109-JUL-1999; 99US-0142920.
112-JUL-1999; 99US-0142977.
113-JUL-1999; 99US-0143542.
114-JUL-1999; 99US-0143624.
115-JUL-1999; 99US-0144005.
116-JUL-1999; 99US-0144085.
116-JUL-1999; 99US-0144086.
119-JUL-1999; 99US-0144325.
119-JUL-1999; 99US-0144331.
```

PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 21-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 11-AUG-1999; 99US-0148171.
PR 12-AUG-1999; 99US-0148319.
PR 13-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 16-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.

PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160983.
PR 22-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 33.9%; Score 40; DB 21; Length 262;
Best Local Similarity 40.0%; Pred. No. 1.4e+02;
Matches 6; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

OY 4 DSCEQQRNVQHYHK 18
| | | | | | | |
Db 7 dgcdsdfsncreykh 21

RESULT 50

AAG50997
ID AAG50997 standard; Protein; 262 AA.

XX AC AAG50997;

XX DT 18-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 64681.

KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.

OS Arabidopsis thaliana.

PN EP1033405-A2.

XX PD 06-SEP-2000.

XX 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.

PR 25-MAR-1999; 99US-0126264.

PR 29-MAR-1999; 99US-0126785.

PR 01-APR-1999; 99US-0127462.

PR 06-APR-1999; 99US-0128234.

PR 08-APR-1999; 99US-0128714.

PR 16-APR-1999; 99US-0129845.

PR 19-APR-1999; 99US-0130077.

PR 21-APR-1999; 99US-0130449.

PR 23-APR-1999; 99US-0130510.

PR 23-APR-1999; 99US-0130891.

Wed Dec 19 08:45:36 2001

us-09-462-625-2_copy_55_75.rag

PR	28-APR-1999;	99US-0131449.	PR	21-JUL-1999;	99US-0144814.
PR	30-APR-1999;	99US-0132048.	PR	21-JUL-1999;	99US-0145086.
PR	30-APR-1999;	99US-0132407.	PR	21-JUL-1999;	99US-0145088.
PR	04-MAY-1999;	99US-0132484.	PR	22-JUL-1999;	99US-0145085.
PR	03-MAY-1999;	99US-0132485.	PR	22-JUL-1999;	99US-0145087.
PR	06-MAY-1999;	99US-0132486.	PR	22-JUL-1999;	99US-0145089.
PR	06-MAY-1999;	99US-0132487.	PR	22-JUL-1999;	99US-0145192.
PR	07-MAY-1999;	99US-0132863.	PR	23-JUL-1999;	99US-0145145.
PR	11-MAY-1999;	99US-0134256.	PR	23-JUL-1999;	99US-0145218.
PR	14-MAY-1999;	99US-0134218.	PR	23-JUL-1999;	99US-0145224.
PR	14-MAY-1999;	99US-0134219.	PR	26-JUL-1999;	99US-0145276.
PR	14-MAY-1999;	99US-0134221.	PR	27-JUL-1999;	99US-0145913.
PR	14-MAY-1999;	99US-0134370.	PR	27-JUL-1999;	99US-0145918.
PR	18-MAY-1999;	99US-0134768.	PR	27-JUL-1999;	99US-0145919.
PR	19-MAY-1999;	99US-0134941.	PR	28-JUL-1999;	99US-0145951.
PR	20-MAY-1999;	99US-0135124.	PR	02-AUG-1999;	99US-0146386.
PR	21-MAY-1999;	99US-0135353.	PR	02-AUG-1999;	99US-0146388.
PR	24-MAY-1999;	99US-0135629.	PR	02-AUG-1999;	99US-0146389.
PR	25-MAY-1999;	99US-0136021.	PR	03-AUG-1999;	99US-0147038.
PR	27-MAY-1999;	99US-0136392.	PR	04-AUG-1999;	99US-0147204.
PR	28-MAY-1999;	99US-0136782.	PR	04-AUG-1999;	99US-0147302.
PR	01-JUN-1999;	99US-0137222.	PR	05-AUG-1999;	99US-0147192.
PR	03-JUN-1999;	99US-0137528.	PR	05-AUG-1999;	99US-0147260.
PR	04-JUN-1999;	99US-0137502.	PR	06-AUG-1999;	99US-0147303.
PR	07-JUN-1999;	99US-0137724.	PR	06-AUG-1999;	99US-0147416.
PR	08-JUN-1999;	99US-0138094.	PR	09-AUG-1999;	99US-0147493.
PR	10-JUN-1999;	99US-0138540.	PR	09-AUG-1999;	99US-0147935.
PR	10-JUN-1999;	99US-0138847.	PR	10-AUG-1999;	99US-0148171.
PR	14-JUN-1999;	99US-0139119.	PR	11-AUG-1999;	99US-0148319.
PR	16-JUN-1999;	99US-0139452.	PR	12-AUG-1999;	99US-0148341.
PR	16-JUN-1999;	99US-0139453.	PR	13-AUG-1999;	99US-0148565.
PR	17-JUN-1999;	99US-0139492.	PR	13-AUG-1999;	99US-0148684.
PR	18-JUN-1999;	99US-0139454.	PR	16-AUG-1999;	99US-0149388.
PR	18-JUN-1999;	99US-0139455.	PR	17-AUG-1999;	99US-0149175.
PR	18-JUN-1999;	99US-0139456.	PR	18-AUG-1999;	99US-0149426.
PR	18-JUN-1999;	99US-0139457.	PR	20-AUG-1999;	99US-0149722.
PR	18-JUN-1999;	99US-0139458.	PR	20-AUG-1999;	99US-0149723.
PR	18-JUN-1999;	99US-0139459.	PR	20-AUG-1999;	99US-0149929.
PR	18-JUN-1999;	99US-0139460.	PR	23-AUG-1999;	99US-0149902.
PR	18-JUN-1999;	99US-0139461.	PR	23-AUG-1999;	99US-0149930.
PR	18-JUN-1999;	99US-0139462.	PR	25-AUG-1999;	99US-0150566.
PR	18-JUN-1999;	99US-0139463.	PR	26-AUG-1999;	99US-0150884.
PR	18-JUN-1999;	99US-0139750.	PR	27-AUG-1999;	99US-0151065.
PR	21-JUN-1999;	99US-0139763.	PR	27-AUG-1999;	99US-0151066.
PR	22-JUN-1999;	99US-0139817.	PR	27-AUG-1999;	99US-0151080.
PR	22-JUN-1999;	99US-0139899.	PR	30-AUG-1999;	99US-0151303.
PR	23-JUN-1999;	99US-0140353.	PR	31-AUG-1999;	99US-0151438.
PR	23-JUN-1999;	99US-0140354.	PR	01-SEP-1999;	99US-0151930.
PR	24-JUN-1999;	99US-0140695.	PR	07-SEP-1999;	99US-0152363.
PR	24-JUN-1999;	99US-0140823.	PR	10-SEP-1999;	99US-0153070.
PR	29-JUN-1999;	99US-0140991.	PR	13-SEP-1999;	99US-0153758.
PR	30-JUN-1999;	99US-0141287.	PR	15-SEP-1999;	99US-0154018.
PR	01-JUL-1999;	99US-0141842.	PR	16-SEP-1999;	99US-0154039.
PR	01-JUL-1999;	99US-0142154.	PR	20-SEP-1999;	99US-0154779.
PR	02-JUL-1999;	99US-0142055.	PR	22-SEP-1999;	99US-0155139.
PR	06-JUL-1999;	99US-0142390.	PR	23-SEP-1999;	99US-0155486.
PR	08-JUL-1999;	99US-0142803.	PR	24-SEP-1999;	99US-0155659.
PR	09-JUL-1999;	99US-0142920.	PR	28-SEP-1999;	99US-0156458.
PR	12-JUL-1999;	99US-0142977.	PR	29-SEP-1999;	99US-0156596.
PR	13-JUL-1999;	99US-0143542.	PR	04-OCT-1999;	99US-0157117.
PR	14-JUL-1999;	99US-0143624.	PR	05-OCT-1999;	99US-0157753.
PR	15-JUL-1999;	99US-0144005.	PR	06-OCT-1999;	99US-0157865.
PR	16-JUL-1999;	99US-0144085.	PR	07-OCT-1999;	99US-0158029.
PR	16-JUL-1999;	99US-0144086.	PR	08-OCT-1999;	99US-0158232.
PR	19-JUL-1999;	99US-0144325.	PR	12-OCT-1999;	99US-0158369.
PR	19-JUL-1999;	99US-0144331.	PR	13-OCT-1999;	99US-0159293.
PR	19-JUL-1999;	99US-0144332.	PR	13-OCT-1999;	99US-0159294.
PR	19-JUL-1999;	99US-0144333.	PR	13-OCT-1999;	99US-0159295.
PR	19-JUL-1999;	99US-0144334.	PR	14-OCT-1999;	99US-0159329.
PR	19-JUL-1999;	99US-0144335.	PR	14-OCT-1999;	99US-0159330.
PR	20-JUL-1999;	99US-0144352.	PR	14-OCT-1999;	99US-0159331.
PR	20-JUL-1999;	99US-0144632.	PR	14-OCT-1999;	99US-0159637.
PR	20-JUL-1999;	99US-0144632.	PR	14-OCT-1999;	99US-0159638.

```
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match      33.9%; Score 40; DB 21; Length 262;
Best Local Similarity 40.0%; Pred. No. 1.4e+02;
Matches 6; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 4 DSCEQCARNVQYHK 18
   | | | | |
Db 7 dgcdsfsncryhk 21
```

Search completed: December 17, 2001, 07:50:51
Job time: 73 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 17, 2001, 09:06:26 ; Search time 164.03 Seconds
(without alignments)
2869.424 Million cell updates/sec

Title: US-09-462-625-1

Perfect score: 549

Sequence: 1 atgtgttgcctgtgctct.....gggaacactaccgagagtga 549

Scoring table: OLIGO_NUC

Gapop 60.0 , Gapext 60.0

Searched: 930621 seqs, 428662619 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : N_Geneseq_1101.*

1: /SID88/gcgdata/geneseq/geneseqn/NA1980.DAT.*
2: /SID88/gcgdata/geneseq/geneseqn/NA1981.DAT.*
3: /SID88/gcgdata/geneseq/geneseqn/NA1982.DAT.*
4: /SID88/gcgdata/geneseq/geneseqn/NA1983.DAT.*
5: /SID88/gcgdata/geneseq/geneseqn/NA1984.DAT.*
6: /SID88/gcgdata/geneseq/geneseqn/NA1985.DAT.*
7: /SID88/gcgdata/geneseq/geneseqn/NA1986.DAT.*
8: /SID88/gcgdata/geneseq/geneseqn/NA1987.DAT.*
9: /SID88/gcgdata/geneseq/geneseqn/NA1988.DAT.*
10: /SID88/gcgdata/geneseq/geneseqn/NA1989.DAT.*
11: /SID88/gcgdata/geneseq/geneseqn/NA1990.DAT.*
12: /SID88/gcgdata/geneseq/geneseqn/NA1991.DAT.*
13: /SID88/gcgdata/geneseq/geneseqn/NA1992.DAT.*
14: /SID88/gcgdata/geneseq/geneseqn/NA1993.DAT.*
15: /SID88/gcgdata/geneseq/geneseqn/NA1994.DAT.*
16: /SID88/gcgdata/geneseq/geneseqn/NA1995.DAT.*
17: /SID88/gcgdata/geneseq/geneseqn/NA1996.DAT.*
18: /SID88/gcgdata/geneseq/geneseqn/NA1997.DAT.*
19: /SID88/gcgdata/geneseq/geneseqn/NA1998.DAT.*
20: /SID88/gcgdata/geneseq/geneseqn/NA1999.DAT.*
21: /SID88/gcgdata/geneseq/geneseqn/NA2000.DAT.*
22: /SID88/gcgdata/geneseq/geneseqn/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	549	100.0	549	AA21819	Mouse tag7 clone c
2	379	69.0	677	AA21819	Murine granulocyte
3	22	4.0	41	AA21827	Primer EBI8203 for
4	22	4.0	41	AA21834	Primer EBI8203 for
5	22	4.0	285	AA80662	Human secreted pro
6	22	4.0	697	AA58104	Human PRO1269 nucl
7	22	4.0	697	AA37082	Human PRO1269 (UNQ
8	22	4.0	697	AA54356	DNA encoding prote
9	22	4.0	718	AA21820	Human tag7 clone c
10	22	4.0	726	AA80613	Human Htag7 secret
11	22	4.0	749	AA51719	Chondrosarcoma pep

12	21	3.8	30	20	AA21828	Primer EBI8471 for
13	19	3.5	3580	22	AAH18181	Human cDNA sequenc
14	18	3.3	18	21	AAC58040	Human PRO1269 for
15	18	3.3	18	22	AAF54520	Primer #130 used i
16	18	3.3	30	20	AA21829	Primer EBI8472 for
17	18	3.3	36	20	AA21826	Primer EBI8025 for
18	18	3.3	36	20	AA21826	Primer EBI8233 for
19	18	3.3	36	20	AA21830	Primer EBI8024 for
20	18	3.3	36	20	AA21825	Primer EBI8024 for
21	18	3.3	362	22	AA211740	Probe #1673 for ge
22	18	3.3	362	22	AAI33047	Probe #1733 used t
23	18	3.3	362	22	AAI01668	Probe #1659 used t
24	18	3.3	675	22	AAE29466	Gerbil TANGO 292 c
25	18	3.3	684	20	AA24239	WO9916900 Seq ID 9
26	18	3.3	684	20	AA24239	WO9916900 Seq ID 9
27	18	3.3	2002	22	AAF29465	Gerbil TANGO 292 c
28	17	3.1	31	20	AA21832	Primer EBI8201 for
29	17	3.1	31	20	AA21833	Primer EBI8202 for
30	17	3.1	359	22	AAH87856	Peppermint plant o
31	17	3.1	688	18	AAH78509	Bovine granulocyte
32	17	3.1	795	20	AA216023	Human gene express
33	17	3.1	1140	22	AAI62821	Human genomic DNA
34	17	3.1	1143	18	AAH58814	Thrombin receptor
35	17	3.1	1146	19	AAV61053	Human clone HACCH9
36	17	3.1	1413	22	AAI62753	Human cDNA SEQ ID
37	17	3.1	1590	22	AAC84864	Chimeric chemokine
38	17	3.1	1630	20	AAH57402	Rat U3 gene trap d
39	17	3.1	1718	22	AAH14807	Human cDNA sequenc
40	17	3.1	2054	22	AAH16990	Human cDNA sequenc
41	17	3.1	2818	22	AAI58215	Human polynucleoti
42	17	3.1	3002	22	AAH63437	Murine ADAMTS-5 cd
43	17	3.1	3133	20	AAH57463	Rat U3 gene trap d
44	17	3.1	3144	21	AAA40083	Human brain-specif
45	17	3.1	3144	21	AAA60605	Human h001149 prot
			3474	22	AAH02894	Human shear stress

ALIGNMENTS

RESULT 1

AA21819
ID AA21819 standard; cDNA; 549 BP.

XX AC AA21819;

XX DT 18-MAY-1999 (first entry)

XX DE Mouse tag7 clone coding sequence.

XX KW Tag7; tumour growth inhibitor; mammalian tumour; carcinoma; sarcoma; melanoma; leukaemia; apoptosis inducer; mouse; ds.

XX OS Mus sp.

XX PN WO9902686-Al.

XX PD 21-JAN-1999.

XX PF 10-JUL-1998; 98WO-EP04287.

XX PR 11-JUL-1997; 97US-0893764.

XX PA (BOEH) BOEHRINGER INGELHEIM INT GMBH.

XX PI Georgiev G, Kiselev S, Ostermann E, Prokhorchouk E;

XX DR WPI; 1999-120887/10.

XX DR P-PSDB; AAY00770.

XX PT New nucleic acid encoding tag7 - used to inhibit tumour growth and induce apoptosis, for treatment of carcinoma, sarcoma, melanoma and leukaemia

QY 181 gaacagcaggccgcaatgtgcagcattaccacaaagaatgagctggctggtgcgatgta 240
 |||||
 Db 216 gaacagcaggccgcaatgtgcagcattaccacaaagaatgagctggctggtgcgatgta 275
 |||||
 QY 241 gctacaacttccttattgagagggcggtcatgtctatgaagccgagcgctggaacatc 300
 |||||
 Db 276 gctacaacttccttattgagagggcggtcatgtctatgaagccgagcgctggaacatc 335
 |||||
 QY 301 aagggtagcacacacagggcccatctggaatcccatctctattgcatccatttcagggg 360
 |||||
 Db 336 aagggtagcacacacagggcccatctggaatcccatctctattgcatccatttcagggg 395
 |||||
 QY 361 aacttcattgacgggtac 379
 |||||
 Db 396 aacttcattgacgggtac 414
 |||||

RESULT 3
 AAX21827/C
 ID AAX21827 standard; DNA; 41 BP.
 XX
 AC AAX21827;
 XX
 DT 18-MAY-1999 (first entry)
 XX
 DE Primer EBI8203 for Mouse tag7 clone coding sequence.
 XX
 KW Tag7; tumour growth inhibitor; mammalian tumour; carcinoma; sarcoma;
 KW melanoma; leukaemia; apoptosis inducer; mouse; PCR primer; ss.
 XX
 OS Synthetic.
 OS Mus sp.
 XX
 PN WO9902686-A1.
 XX
 PD 21-JAN-1999.
 XX
 PF 10-JUL-1998; 98WO-EP04287.
 XX
 PR 11-JUL-1997; 97US-0893764.
 XX
 PA (BOEH) BOEHRINGER INGELHEIM INT GMBH.
 XX
 PI Georgiev G, Kiselev S, Ostermann E, Prokhorchouk E;
 XX
 DR WPI; 1999-120887/10.
 XX
 PS New nucleic acid encoding tag7 - used to inhibit tumour growth and
 PT induce apoptosis, for treatment of carcinoma, sarcoma, melanoma and
 PT leukaemia
 XX
 PS Example 13; Page 74; 138pp; English.

XX This sequence is a PCR primer for DNA encoding the murine tag7 of the
 CC invention. Cells containing the tag7 DNA sequence are used to express
 CC recombinant tag7. Tag7 is used to produce and purify antibodies; to
 CC inhibit growth of mammalian tumours, especially for treating carcinoma
 CC (of liver, ovary, breast, cervix, lung, prostate, colon/rectum, bladder,
 CC testis, stomach, pancreas, mouth, head and neck, squamous cell carcinoma
 CC or teratocarcinoma), sarcoma (Kaposi's, osteo- or fibro-sarcomas),
 CC melanoma or leukaemia; and as a molecular weight marker. The tag7
 CC polypeptide inhibits tumour growth and induces apoptosis. The tag7
 CC sequences are also useful as probes for gene mapping and detection of
 CC tag7 gene expression, and as primers. Antibodies against tag7 are used as
 CC reagents for detecting tag7; as an antagonist of tag7; for isolating tag7
 CC and therapeutically to inhibit or delay tumour metastasis.
 XX

Sequence 41 BP; 6 A; 13 C; 11 G; 11 T; 0 other;

Query Match 4.0%; Score 22; DB 20; Length 41;
 Best Local Similarity 100.0%; Pred. No. 0.21;

Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 528 ctgggaacactaccgagagtga 549
 |||||
 Db 41 CTGGGAACACTACCAGAGAGTGA 20

RESULT 4
 AAX21834/C
 ID AAX21834 standard; DNA; 41 BP.
 XX
 AC AAX21834;
 XX

DT 18-MAY-1999 (first entry)
 XX
 DE Primer EBI8203 for Mouse tag7 clone coding sequence.
 XX

KW Tag7; tumour growth inhibitor; mammalian tumour; carcinoma; sarcoma;
 KW melanoma; leukaemia; apoptosis inducer; mouse; PCR primer; ss.

OS Synthetic.
 OS Mus sp.

PN WO9902686-A1.

XX 21-JAN-1999.

XX 10-JUL-1998; 98WO-EP04287.

XX 11-JUL-1997; 97US-0893764.

XX (BOEH) BOEHRINGER INGELHEIM INT GMBH.

XX Georgiev G, Kiselev S, Ostermann E, Prokhorchouk E;

XX WPI; 1999-120887/10.

XX New nucleic acid encoding tag7 - used to inhibit tumour growth and
 PT induce apoptosis, for treatment of carcinoma, sarcoma, melanoma and
 PT leukaemia
 XX

XX Example 13; Page 77; 138pp; English.

XX This sequence is a PCR primer for DNA encoding the murine tag7 of the
 CC invention. Cells containing the tag7 DNA sequence are used to express
 CC recombinant tag7. Tag7 is used to produce and purify antibodies; to
 CC inhibit growth of mammalian tumours, especially for treating carcinoma
 CC (of liver, ovary, breast, cervix, lung, prostate, colon/rectum, bladder,
 CC testis, stomach, pancreas, mouth, head and neck, squamous cell carcinoma
 CC or teratocarcinoma), sarcoma (Kaposi's, osteo- or fibro-sarcomas),
 CC melanoma or leukaemia; and as a molecular weight marker. The tag7
 CC polypeptide inhibits tumour growth and induces apoptosis. The tag7
 CC sequences are also useful as probes for gene mapping and detection of
 CC tag7 gene expression, and as primers. Antibodies against tag7 are used as
 CC reagents for detecting tag7; as an antagonist of tag7; for isolating tag7
 CC and therapeutically to inhibit or delay tumour metastasis.
 XX

Sequence 41 BP; 6 A; 13 C; 11 G; 11 T; 0 other;

Query Match 4.0%; Score 22; DB 20; Length 41;
 Best Local Similarity 100.0%; Pred. No. 0.21;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 528 ctgggaacactaccgagagtga 549
 |||||
 Db 41 CTGGGAACACTACCAGAGAGTGA 20

RESULT 5
 AAA80662
 ID AAA80662 standard; CDNA; 285 BP.
 XX

XX	RESULT	6
XX	AAC58104	
ID	AAC58104 standard; cDNA; 697 BP.	
XX	AAC58104;	
XX	AAC	
XX	25-JAN-2001 (first entry)	
XX	Human PRO1269 nucleotide sequence SEQ ID NO:6.	
XX	Human; tumour; diagnosis; neoplastic disease; proliferation; cancer;	
KW	identification; tumourigenesis; anticancer; detection; ss.	
XX	Homo sapiens.	
XX	WO200053750-A1.	
XX	14-SEP-2000.	
XX	02-DEC-1999; 99WO-US28551.	
XX	08-MAR-1999; 99WO-US05028.	
PR	01-SEP-1999; 99WO-US20111.	
PR	29-OCT-1999; 99US-0162506.	
PR	30-NOV-1999; 99WO-US28313.	
PR	01-DEC-1999; 99WO-US28634.	
XX	(GETH) GENENTECH INC.	
XX	Botstein D, Goddard A, Gurney AL, Roy MA, Watanabe CK, Wood WI;	
PI	WPI; 2000-594320/56.	
DR	P-PSDB; AAB24022.	
XX	Antibodies specific for PRO polypeptides, used to diagnose and inhibit	
PT	the growth of tumors in mammals, and to identify inhibitors of PRO	
PT	polypeptide activity or expression	
XX	Claim 50; Fig 3; 226pp; English.	
PS	The present invention describes an antibody that binds to a human	
XX	protein (I) selected from: PRO381; PRO1269; PRO1410; PRO1755; PRO1780;	
CC	PRO3434; PRO1927; PRO3567; PRO1295; PRO1303; PRO4344; PRO4354;	
CC	PRO4397; PRO4407; PRO1555; PRO1096; PRO2038; and PRO2262. (I) has	
CC	anticancer activity and can be used to diagnose tumours in mammals, by	
CC	detecting complex formation when the antibody is contacted with test	
CC	cells. Increased expression of genes encoding (I) can also be detected	
CC	to diagnose tumours. Agents which inhibit the activity of (I),	
CC	especially the antibodies, or an antisense oligonucleotide which	
CC	hybridises to genes encoding (I), can be used to inhibit tumour growth,	
CC	preferably by inducing cell death. Methods from the present invention	
CC	can be used to identify compounds which inhibit the biological activity	
CC	of (I). AAC58019 to AAC58102 represent PCR primers and hybridisation	
CC	probes used in examples from the present invention for human PRO	
CC	sequences. AAC58103 to AAC58122 and AAB24021 to AAB24040 represent human	
CC	PRO polynucleotide and protein sequences given in the exemplification of	
CC	the present invention.	
XX	Sequence 697 BP; 136 A; 250 C; 184 G; 127 T; 0 other;	
XX		

Query Match	4.0%;	Score 22;	DB 21;	Length 697;
Best Local Similarity	100.0%;	Prod. No. 0.19;		
Matches 22;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	461	tcaaaggacacgggatgtgca	482	
Db	525	tcaaaggacacgggatgtgca	546	
RESULT	7			
AAA37082				
ID AAA37082	standard;	cdna;	697 BP.	

XX
AC AAA37082;
XX
DT 08-AUG-2000 (first entry)
XX
DE Human PRO1269 (UNQ639) cDNA sequence SEQ ID NO:215.
XX
KW Human; PRO polypeptide; membrane bound protein; receptor; diagnosis;
KW transmembrane; secretion; immunoadhesion; pharmaceutical; screening;
KW ss.
XX
OS Homo sapiens.
XX
PN WO200012708-A2.
XX
PD 09-MAR-2000.
XX
PF 01-SEP-1999; 99WO-US20111.
XX
PR 01-SEP-1998; 98US-0098716.
PR 01-SEP-1998; 98US-0098749.
PR 01-SEP-1998; 98US-0098750.
PR 02-SEP-1998; 98US-0098803.
PR 02-SEP-1998; 98US-0098821.
PR 02-SEP-1998; 98US-0098843.
PR 02-SEP-1998; 98US-0098936.
PR 02-SEP-1998; 98US-0098956.
PR 09-SEP-1998; 98US-0099598.
PR 09-SEP-1998; 98US-0099602.
PR 09-SEP-1998; 98US-0099642.
PR 10-SEP-1998; 98US-0099741.
PR 10-SEP-1998; 98US-0099754.
PR 10-SEP-1998; 98US-0099763.
PR 10-SEP-1998; 98US-0099792.
PR 10-SEP-1998; 98US-0099808.
PR 10-SEP-1998; 98US-0099812.
PR 10-SEP-1998; 98US-0099815.
PR 10-SEP-1998; 98US-0099816.
PR 15-SEP-1998; 98US-0100385.
PR 15-SEP-1998; 98US-0100388.
PR 15-SEP-1998; 98US-0100390.
PR 16-SEP-1998; 98US-0100584.
PR 16-SEP-1998; 98US-0100627.
PR 16-SEP-1998; 98US-0100661.
PR 16-SEP-1998; 98US-0100662.
PR 16-SEP-1998; 98US-0100664.
PR 17-SEP-1998; 98US-0100683.
PR 17-SEP-1998; 98US-0100710.
PR 17-SEP-1998; 98US-0100711.
PR 17-SEP-1998; 98US-0100919.
PR 17-SEP-1998; 98US-0100930.
PR 18-SEP-1998; 98US-0100848.
PR 18-SEP-1998; 98US-0100849.
PR 18-SEP-1998; 98US-0101014.
PR 18-SEP-1998; 98US-0101068.
PR 18-SEP-1998; 98US-0101071.
PR 22-SEP-1998; 98US-0101279.
PR 23-SEP-1998; 98US-0101471.
PR 23-SEP-1998; 98US-0101472.
PR 23-SEP-1998; 98US-0101474.
PR 23-SEP-1998; 98US-0101475.
PR 23-SEP-1998; 98US-0101476.
PR 23-SEP-1998; 98US-0101477.
PR 23-SEP-1998; 98US-0101479.
PR 24-SEP-1998; 98US-0101738.
PR 24-SEP-1998; 98US-0101741.
PR 24-SEP-1998; 98US-0101743.
PR 24-SEP-1998; 98US-0101915.
PR 24-SEP-1998; 98US-0101916.
PR 29-SEP-1998; 98US-0102207.
PR 29-SEP-1998; 98US-0102240.
PR 29-SEP-1998; 98US-0102307.

PR 29-SEP-1998; 98US-0102330.
PR 29-SEP-1998; 98US-0102331.
PR 30-SEP-1998; 98US-0102484.
PR 30-SEP-1998; 98US-0102487.
PR 30-SEP-1998; 98US-0102570.
PR 30-SEP-1998; 98US-0102571.
PR 01-OCT-1998; 98US-0102684.
PR 01-OCT-1998; 98US-0102687.
PR 02-OCT-1998; 98US-0102965.
PR 06-OCT-1998; 98US-0103258.
PR 06-OCT-1998; 98US-0103449.
PR 06-OCT-1998; 98US-0103314.
PR 07-OCT-1998; 98US-0103315.
PR 07-OCT-1998; 98US-0103328.
PR 07-OCT-1998; 98US-0103395.
PR 07-OCT-1998; 98US-0103396.
PR 07-OCT-1998; 98US-0103401.
PR 08-OCT-1998; 98US-0103633.
PR 08-OCT-1998; 98US-0103678.
PR 08-OCT-1998; 98US-0103679.
PR 08-OCT-1998; 98US-0103711.
PR 14-OCT-1998; 98US-0104257.
PR 20-OCT-1998; 98US-0104987.
PR 20-OCT-1998; 98US-0105000.
PR 20-OCT-1998; 98US-0105002.
PR 21-OCT-1998; 98US-0105104.
PR 22-OCT-1998; 98US-0105169.
PR 22-OCT-1998; 98US-0105266.
PR 26-OCT-1998; 98US-0105693.
PR 26-OCT-1998; 98US-0105694.
PR 26-OCT-1998; 98US-0105807.
PR 27-OCT-1998; 98US-0105881.
PR 27-OCT-1998; 98US-0105882.
PR 27-OCT-1998; 98US-0106023.
PR 28-OCT-1998; 98US-0106029.
PR 28-OCT-1998; 98US-0106030.
PR 28-OCT-1998; 98US-0106032.
PR 28-OCT-1998; 98US-0106033.
PR 28-OCT-1998; 98US-0106178.
PR 29-OCT-1998; 98US-0106248.
PR 29-OCT-1998; 98US-0106384.
PR 29-OCT-1998; 98US-0108500.
PR 30-OCT-1998; 98US-0108464.
PR 03-NOV-1998; 98US-0106856.
PR 03-NOV-1998; 98US-0106902.
PR 03-NOV-1998; 98US-0106905.
PR 03-NOV-1998; 98US-0106919.
PR 03-NOV-1998; 98US-0106932.
PR 03-NOV-1998; 98US-0106934.
PR 10-NOV-1998; 98US-0107783.
PR 10-NOV-1998; 98US-0108775.
PR 17-NOV-1998; 98US-0108779.
PR 17-NOV-1998; 98US-0108787.
PR 17-NOV-1998; 98US-0108788.
PR 17-NOV-1998; 98US-0108801.
PR 17-NOV-1998; 98US-0108802.
PR 17-NOV-1998; 98US-0108806.
PR 17-NOV-1998; 98US-0108807.
PR 17-NOV-1998; 98US-0108867.
PR 17-NOV-1998; 98US-0108925.
PR 18-NOV-1998; 98US-0108848.
PR 18-NOV-1998; 98US-0108849.
PR 18-NOV-1998; 98US-0108850.
PR 18-NOV-1998; 98US-0108851.
PR 18-NOV-1998; 98US-0108852.
PR 18-NOV-1998; 98US-0108858.
PR 18-NOV-1998; 98US-0108904.

(GETH) GENENTECH INC.

Baker K, Goddard A, Gurney AL, Smith V, Watanabe CK, Wood WI;

XX
XX
PI
XX

Wed Dec 19 08:45:12 2001

us-09-462-625-1.oli.rng

DR WPI: 2000-237871/20.
 DR P-PSDB; AAY99400.
 XX
 XX
 PT New mammalian DNA sequences encoding transmembrane, receptor or
 PT secreted PRO polypeptides, useful for screening of potential peptide or
 PT small molecule inhibitors of the relevant receptor/ligand interactions
 XX
 XX
 PS Claim 2; Fig 121; 773pp; English.
 XX
 XX AAA37022 to AAA37144 encode the new isolated human transmembrane,
 CC receptor or secreted PRO polypeptides given in AAY99340 to AAY99462. The
 CC transmembrane and receptor PRO proteins can be used for screening of
 CC potential peptide or small molecule inhibitors of the relevant
 CC receptor/ligand interactions. The polypeptides and nucleotide sequences
 CC encoding then have various industrial applications, including uses as
 CC pharmaceutical and diagnostic agents. AAA37145 to AAA37330 represent
 CC PCR primers and hybridisation probes used in the isolation of the PRO
 CC polypeptides from the present invention.
 XX
 SQ Sequence 697 BP; 136 A; 250 C; 184 G; 127 T; 0 other;
 Query Match 4.0%; Score 22; DB 21; Length 697;
 Best Local Similarity 100.0%; Pred. No. 0.19;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 461 tcaaaaggacacgggatgtgca 482
 Db 525 tcaaaaggacacgggatgtgca 546
 RESULT 8
 AAF54356
 ID AAF54356 standard; DNA; 697 BP.
 XX
 AC AAF54356;
 DT 02-APR-2001 (first entry)
 XX
 XX DNA encoding protein of the invention #61.
 DE
 XX Secreted; transmembrane; gene therapy; ss.
 XX
 OS Unidentified.
 XX
 PN WO200078961-A1.
 XX
 PD 28-DEC-2000.
 XX
 PF 18-FEB-2000; 2000WO-US04342.
 XX
 XX 23-JUN-1999; 99US-0141037.
 PR 20-JUL-1999; 99US-0144758.
 PR 26-JUL-1999; 99US-0145698.
 PR 01-SEP-1999; 99WO-US20111.
 PR 29-OCT-1999; 99US-0162506.
 PR 30-NOV-1999; 99WO-US28313.
 PR 02-DEC-1999; 99WO-US28551.
 PR 16-DEC-1999; 99WO-US30095.
 PR 05-JAN-2000; 2000WO-US00219.
 PR 06-JAN-2000; 2000WO-US00376.
 XX
 XX (GETH) GENENTECH INC.
 XX Baker KP, Botstein D, Desnoyers L, Eaton DL, Ferrara N, Fong S;
 PI Gao W, Goddard A, Godowski PJ, Grimaldi CJ, Gurney AL, Hillan KJ;
 PI Pan J, Paoni NF, Roy MA, Smith V, Stewart TA, Tumas D;
 PI Watanabe CK, Williams PM, Wood WI;
 XX WPI: 2001-071395/08.
 XX
 XX Secreted and transmembrane proteins and nucleic acids designated PRO,
 PT useful as hybridization probes, in chromosome and gene mapping and gene

therapy -
 PT
 XX Claim 2; Fig 121; 787pp; English.
 XX
 CC The present invention relates to secreted and transmembrane proteins.
 CC These proteins and the DNA encoding them may be used as hybridization
 CC probes, in chromosome and gene mapping and in the generation of
 CC anti-sense RNA and DNA. They may also be used to generate either
 CC transgenic animals or knockout animals which are in turn useful for
 CC development and screening of therapeutically useful reagents.
 CC The nucleic acids may also be used in gene therapy.
 XX
 SQ Sequence 697 BP; 136 A; 250 C; 184 G; 127 T; 0 other;
 Query Match 4.0%; Score 22; DB 22; Length 697;
 Best Local Similarity 100.0%; Pred. No. 0.19;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 461 tcaaaaggacacgggatgtgca 482
 Db 525 tcaaaaggacacgggatgtgca 546
 RESULT 9
 AAX21820
 ID AAX21820 standard; cDNato mRNA; 718 BP.
 XX
 AC AAX21820;
 XX
 DT 18-MAY-1999 (first entry)
 XX
 DE Human tag7 clone coding sequence.
 XX
 XX Tag7; tumour growth inhibitor; mammalian tumour; carcinoma; sarcoma;
 KW melanoma; leukaemia; apoptosis inducer; human; ss.
 KW
 XX Homo sapiens.
 OS
 XX WO9902686-A1.
 PN
 XX 21-JAN-1999.
 PD
 XX 10-JUL-1998; 98WO-EP04287.
 PP
 XX 11-JUL-1997; 97US-0893764.
 PR
 XX (BOEH) BOEHRINGER INGELHEIM INT GMBH.
 PA
 XX Georgiev G, Kiselev S, Ostermann E, Prokhorchouk E;
 PI
 DR WPI: 1999-120887/10.
 DR P-PSDB; AAY00771.
 XX
 XX New nucleic acid encoding tag7 - used to inhibit tumour growth and
 PT induce apoptosis, for treatment of carcinoma, sarcoma, melanoma and
 PT leukaemia
 XX
 PS Claim 11; Page 126-127; 138pp; English.
 XX
 XX This sequence encodes the human tag7 of the invention. Cells containing
 CC the tag7 DNA sequence are used to express recombinant tag7. tag7 is used
 CC to produce and purify antibodies; to inhibit growth of mammalian tumours,
 CC especially for treating carcinoma (of liver, ovary, breast, cervix, lung,
 CC prostate, colon/rectum, bladder, testis, stomach, pancreas, mouth, head
 CC and neck, squamous cell carcinoma or teratocarcinoma), sarcoma (Kaposi's,
 CC osteo- or fibro-sarcomas), melanoma or leukaemia; and as a molecular
 CC weight marker. The tag7 polypeptide inhibits tumour growth and induces
 CC apoptosis. The tag7 coding sequences are also useful as probes for gene
 CC mapping and detection of tag7 gene expression, and as primers. Antibodies
 CC against tag7 are used as reagents for detecting tag7; as an antagonist of
 CC tag7; for isolating tag7 and therapeutically to inhibit or delay tumour
 CC metastasis.

XX SQ Sequence 718 BP; 142 A; 251 C; 194 G; 131 T; 0 other;

Query Match 4.0%; Score 22; DB 20; Length 718;
Best Local Similarity 100.0%; Pred. No. 0.19;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 461 tcaaaaggacacccgggatgtgca 482
|||||
Db 552 tcaaaaggacacccgggatgtgca 573
|||||

RESULT 10
AA80613
ID AAA80613 standard; cDNA; 726 BP.
XX
AC AA80613;
XX
DT 21-NOV-2000 (first entry)
XX
DE Human Htag7 secreted protein gene #8.
XX
KW Secreted protein; immunosuppressant; anti-inflammatory; antiarthritic;
KW antirheumatic, dermatological; antiproliferative; antiarteriosclerotic;
KW anticancer; vulnary; antiviral; antibacterial; antifungal;
KW immune disorder; Addison's disease; rheumatoid arthritis; dermatitis;
KW multiple sclerosis; inflammatory disorder; inflammatory bowel disease;
KW Crohn's disease; nephritis; hyperproliferative disorder;
KW cardiovascular disorder; coronary arteriosclerosis; myocarditis; cancer;
KW melanoma; lymphoma; wound healing; human; ss.
XX
OS Homo sapiens.
XX
PN WO200029435-A1.
XX
PD 25-MAY-2000.
XX
PF 27-OCT-1999; 99WO-US25031.
XX
PR 28-OCT-1998; 98US-0105971.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Ni J. Ruben SM, Olsen HS, Young PE, Kenny JJ, Moore PA, Wei Y;
PI Greene JM;
XX
DR WPI; 2000-387742/33.
XX
DR P-PSDB; AAB25583.
XX
XX Isolated nucleic acid molecules encoding human secreted proteins are
XX used for the prevention, amelioration and treatment of autoimmune,
XX inflammatory, hyperproliferative and cardiovascular disorders, cancer,
XX wounds, and infectious diseases -
XX
XX Claim 1; Figure 34; 803pp; English.

CC The present invention relates to 12 secreted human proteins and the
CC nucleotide sequences encoding them. The polynucleotide sequences given
CC in AAA80606-A80623 encode the 12 secreted protein sequences given in
CC AAB25576-B25593. The human secreted proteins have various activities
CC dependent on the tissues in which they are expressed. Examples of the
CC activities of the proteins include: immunosuppressant;
CC anti-inflammatory; antiarthritic; antirheumatic, dermatological;
CC antiproliferative; antiarteriosclerotic; anticancer; vulnary;
CC antiviral; antibacterial; and antifungal activity. The proteins,
CC polypeptides, agonists and antagonists may be used to treat prevent
CC and/or diagnose various disease, disorders and conditions examples of
CC which include: immune disorders e.g. Addison's disease, rheumatoid
CC arthritis, dermatitis, and multiple sclerosis; inflammatory disorders
CC e.g. inflammatory bowel disease, Crohn's disease and nephritis;
CC hyperproliferative disorders such as paraproteinemias and purpura;
CC cardiovascular disorders e.g. coronary arteriosclerosis and myocarditis;

CC cancer e.g. melanoma and lymphoma. The proteins and polynucleotide
CC sequences may also be used in wound healing and the treatment of
CC infectious diseases. The human secreted protein gene #8 and protein
CC sequences are represented in sequences AAA80613 and AAB25583. Sequences
CC AAA80662-A80663 represent genes related to the secreted protein gene#8.
XX
SQ Sequence 726 BP; 161 A; 251 C; 188 G; 126 T; 0 other;

Query Match 4.0%; Score 22; DB 21; Length 726;
Best Local Similarity 100.0%; Pred. No. 0.19;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 461 tcaaaaggacacccgggatgtgca 482
|||||
Db 531 tcaaaaggacacccgggatgtgca 552
|||||

RESULT 11
AA51719
ID AA51719 standard; cDNA; 749 BP.
XX
AC AA51719;
XX
DT 31-OCT-2000 (first entry)
XX
DE Chondrosarcoma peptidoglycan recognition protein-like protein cDNA.
XX
KW Peptidoglycan recognition protein-like protein; PGRP-C; regulator;
KW chondrosarcoma; immunosuppressive; antibacterial; anti-apoptotic;
KW tumour necrosis factor; TNF; anti-infectious; mononuclear phagocyte;
KW inhibitor; protein co-ordinate data; ss.
XX
OS Homo sapiens.
XX
FH Key
CDS 55..645
FT FT /*tag= a
FT FT /product= PGRP-like_protein
FT FT 55..117
FT FT /*tag= b
FT FT 118..642
FT FT /*tag= c

XX WO200039327-A1.
XX
PN PN
XX
PD 06-JUL-2000.
XX
PF 22-DEC-1999; 99WO-US30736.
XX
PR 23-DEC-1998; 98US-0113809.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Ruben SM, Young PE, Olsen HS;
XX
XX WPI; 2000-452414/39.
XX
DR P-PSDB; AAY96964.
XX
XX Polynucleotide encoding peptidoglycan recognition protein-like protein,
XX antibodies specific to it useful for preventing, treating conditions
XX e.g. endotoxic shock and auto-immune disorders and infections in mammal
XX
XX Claim 1; Fig 3; 191pp; English.

CC Novel human peptidoglycan recognition protein-like proteins (PGRP)
CC expressed by keratinocytes (PGRP-K), wound healing tissues (PGRP-W)
CC or chondrosarcoma tissue (PGRP-C) are disclosed. The PGRP DNA and
CC proteins are useful for preventing, treating or ameliorating a medical
CC condition in a mammal (claimed). PGRP is useful in augmenting the immune
CC system in such areas as immune recognition, antigen presentation and
CC immune system activation. Antibodies or antagonists directed against
CC these proteins may be useful in reducing or eliminating disorders

Wed Dec 19 08:45:12 2001

us-09-462-625-1.oli.rng

CC associated with tumour necrosis factor (TNF) and TNF-like cytokines,
CC such as endotoxic shock and autoimmune disorders and for treating
CC infectious diseases including silicosis, sarcoidosis and idiopathic
CC pulmonary fibrosis.

XX Sequence 749 BP; 165 A; 259 C; 195 G; 130 T; 0 other;

Query Match 4.0%; Score 22; DB 21; Length 749;
Best Local Similarity 100.0%; Pred. No. 0.19;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 461 tcaaaaggacaccgggatgtgca 482
|||||
Db 554 tcaaaaggacaccgggatgtgca 575

RESULT 12

AAX21828
ID AAX21828 standard; DNA; 30 BP.

AC AAX21828;

XX 18-MAY-1999 (first entry)

DT Primer EBI8471 for Mouse tag7 clone coding sequence.

DE Tag7: tumour growth inhibitor; mammalian tumour; carcinoma; sarcoma;
KW melanoma; leukaemia; apoptosis inducer; mouse; PCR primer; ss.

XX Synthetic.

OS Mus sp.

XX WO9902686-A1.

XX 21-JAN-1999.

XX 10-JUL-1998; 98WO-EP04287.

XX 11-JUL-1997; 97US-0893764.

XX (BOEHR) BOEHRINGER INGELHEIM INT GMBH.

XX Georgiev G, Kiselev S, Ostermann E, Prokhorchouk E;

XX WPI; 1999-120887/10.

XX New nucleic acid encoding tag7 - used to inhibit tumour growth and
PT induce apoptosis, for treatment of carcinoma, sarcoma, melanoma and
PT leukaemia

XX Example 13; Page 75; 138pp; English.

XX This sequence is a PCR primer for DNA encoding the murine tag7 of the
CC invention. Cells containing the tag7 DNA sequence are used to express
CC recombinant tag7. Tag7 is used to produce and purify antibodies; to
CC inhibit growth of mammalian tumours, especially for treating carcinoma
CC (of liver, ovary, breast, cervix, lung, prostate, colon/rectum, bladder,
CC testis, stomach, pancreas, mouth, head and neck, squamous cell carcinoma
CC or teratocarcinoma), sarcoma (Kaposi's, osteo- or fibro-sarcomas),
CC melanoma or leukaemia; and as a molecular weight marker. The tag7
CC polypeptide inhibits tumour growth and induces apoptosis. The tag7 coding
CC sequences are also useful as probes for gene mapping and detection of
CC tag7 gene expression, and as primers. Antibodies against tag7 are used as
CC reagents for detecting tag7; as an antagonist of tag7; for isolating tag7
CC and therapeutically to inhibit or delay tumour metastasis.

XX Sequence 30 BP; 4 A; 10 C; 12 G; 4 T; 0 other;

Query Match 3.8%; Score 21; DB 20; Length 30;
Best Local Similarity 100.0%; Pred. No. 0.68;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 73 gagtggaggccctgcaccc 93
|||||
Db 10 gagtggaggccctgcaccc 30

RESULT 13

AAH18181
ID AAH18181 standard; cDNA; 3580 BP.

XX AAH18181;

XX 26-JUN-2001 (first entry)

XX Human cDNA sequence SEQ ID NO:18086.

XX Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.

XX Homo sapiens.

XX EPI074617-A2.

XX 07-FEB-2001.

XX 28-JUL-2000; 2000EP-0116126.

XX 29-JUL-1999; 99JP-0248036.

XX 27-AUG-1999; 99JP-0300253.

XX 11-JAN-2000; 2000JP-0118776.

XX 02-MAY-2000; 2000JP-0183767.

XX 09-JUN-2000; 2000JP-0241899.

XX (HELI-) HELIX RES INST.

XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

XX WPI; 2001-318749/34.

XX Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -

XX Claim 8; SEQ ID 18086; 2537pp + CD ROM; English.

XX The present invention describes primer sets for synthesising 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesising polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.

XX Sequence 3580 BP; 917 A; 909 C; 949 G; 805 T; 0 other;

Query Match 3.5%; Score 19; DB 22; Length 3580;

Best Local Similarity 100.0%; Pred. No. 5.9; Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 25 gccctcctgggtctggcaa 43
|||||
Db 1062 gccctcctgggtctggcaa 1080

RESULT 14
AAC58040
ID AAC58040 standard; DNA; 18 BP.
XX
AC AAC58040;
XX
DT 25-JAN-2001 (first entry)
XX
DE Human PRO1269 forward PCR primer SEQ ID NO:62.
XX
KW Human; tumour; diagnosis; neoplastic disease; proliferation; cancer;
KW identification; tumourigenesis; anticancer; detection; hybridisation;
KW probe; PCR primer; ss.
XX
OS Homo sapiens.
XX
PN WO200053750-A1.
XX
PD 14-SEP-2000.
XX
PF 02-DEC-1999; 99WO-US28551.
XX
PR 08-MAR-1999; 99WO-US05028.
PR 01-SEP-1999; 99WO-US20111.
PR 29-OCT-1999; 99US-0162506.
PR 30-NOV-1999; 99WO-US28313.
PR 01-DEC-1999; 99WO-US28634.
XX
PA (GETH) GENENTECH INC.
XX
PI Botstein D, Goddard A, Gurney AL, Roy MA, Watanabe CK, Wood WI;
XX
DR WPI; 2000-594320/56.
XX
PT Antibodies specific for PRO polypeptides, used to diagnose and inhibit
PT the growth of tumors in mammals, and to identify inhibitors of PRO
PT polypeptide activity or expression -
XX
PS Example 20; Page 122; 226pp; English.
XX
CC The present invention describes an antibody that binds to a human
CC protein (I) selected from: PRO381; PRO1269; PRO1410; PRO1755; PRO1780;
CC PRO3434; PRO1927; PRO3567; PRO1295; PRO1293; PRO1303; PRO4344; PRO4354;
CC PRO4397; PRO4407; PRO1555; PRO1096; PRO2038; and PRO2262. (I) has
CC anticancer activity and can be used to diagnose tumours in mammals, by
CC detecting complex formation when the antibody is contacted with test
CC cells. Increased expression of genes encoding (I) can also be detected
CC to diagnose tumours. Agents which inhibit the activity of (I),
CC especially the antibodies, or an antisense oligonucleotide which
CC hybridises to genes encoding (I), can be used to inhibit tumour growth,
CC preferably by inducing cell death. Methods from the present invention
CC can be used to identify compounds which inhibit the biological activity
CC of (I). AAC58019 to AAC58102 represent PCR primers and hybridisation
CC probes used in examples from the present invention for human PRO
CC sequences. AAC58103 to AAC58122 and AAB24021 to AAB24040 represent human
CC PRO polynucleotide and protein sequences given in the exemplification of
CC the present invention.
XX
SQ Sequence 18 BP; 6 A; 3 C; 7 G; 2 T; 0 other;

Query Match 3.3%; Score 18; DB 21; Length 18;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 463 aaaggacaccgggatgtg 480
|||||
Db 1 aaaggacaccgggatgtg 18
RESULT 15
AAF54520
ID AAF54520 standard; DNA; 18 BP.
XX
AC AAF54520;
XX
DT 02-APR-2001 (first entry)
XX
DE Primer #130 used in the identification of proteins.
XX
KW Secreted; transmembrane; gene therapy; ss.
XX
OS Unidentified.
XX
PN WO200078961-A1.
XX
PD 28-DEC-2000.
XX
PF 18-FEB-2000; 2000WO-US04342.
XX
PR 23-JUN-1999; 99US-0141037.
PR 20-JUL-1999; 99US-0144758.
PR 26-JUL-1999; 99US-0145698.
PR 01-SEP-1999; 99WO-US20111.
PR 29-OCT-1999; 99US-0162506.
PR 30-NOV-1999; 99WO-US28313.
PR 02-DEC-1999; 99WO-US28551.
PR 16-DEC-1999; 99WO-US30095.
PR 05-JAN-2000; 2000WO-US00219.
PR 06-JAN-2000; 2000WO-US00376.
XX
PA (GETH) GENENTECH INC.
XX
PI Baker KP, Botstein D, Desnoyers L, Eaton DL, Ferrara N, Fong S;
PI Gao W, Goddard A, Godowski PJ, Grimaldi CJ, Gurney AL, Hillan KJ;
PI Pan J, Paoni NF, Roy MA, Smith V, Stewart TA, Tumas D;
PI Watanabe CK, Williams PM, Wood WI;
XX
DR WPI; 2001-071395/08.
XX
PT Secreted and transmembrane proteins and nucleic acids designated PRO,
PT useful as hybridization probes, in chromosome and gene mapping and gene
PT therapy -
XX
PS Example 143; Page 507; 787pp; English.
XX
CC The present invention relates to secreted and transmembrane proteins.
CC These proteins and the DNA encoding them may be used as hybridization
CC probes, in chromosome and gene mapping and in the generation of
CC anti-sense RNA and DNA. They may also be used to generate either
CC transgenic animals or knockout animals which are in turn useful for
CC development and screening of therapeutically useful reagents.
CC The nucleic acids may also be used in gene therapy.
XX
SQ Sequence 18 BP; 6 A; 3 C; 7 G; 2 T; 0 other;

Query Match 3.3%; Score 18; DB 22; Length 18;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 463 aaaggacaccgggatgtg 480
|||||
Db 1 aaaggacaccgggatgtg 18

Wed Dec 19 08:45:12 2001

us-09-462-625-1.oli.rng

Page 10

Job time: 5051 sec

```

/cores/segment/segment1
seq name: /cores/condata/geneseq/geneseq/NA1999.DAT.AAX21819

```

seq_documentation_block:
ID AAX21819 standard: cDNA: 549 BP.

seq_documentation_block:
ID AAX21819 standard: cDNA: 549 BP.

Search information blo
Over: HC-08-467-625-2

Query length: 182

Query length: 102
Database: N Geneset 1101.**

Database sequences: 930621

Search time (sec): 166.110000

score_list:	Sequence	Strd	Orig	zScore	EScore	Documentation
	/SID58/cgcgdata/geneseq/geneseq/NA1999.DAT.AAX21819 +			996.00	2028.98	
	/SID58/cgcgdata/geneseq/geneseq/NA1997.DAT.AAX78510 +			944.50	1920.88	
	/SID58/cgcgdata/geneseq/geneseq/NA2000.DAT.AAC58104 +			692.00	1401.67	
	/SID58/cgcgdata/geneseq/geneseq/NA2000.DAT.AAC37082 +			692.00	1401.67	
	/SID58/cgcgdata/geneseq/geneseq/NA2001.DAT.AAP54356 +			692.00	1401.67	
	/SID58/cgcgdata/geneseq/geneseq/NA1999.DAT.AAX21820 +			692.00	1401.33	
	/SID58/cgcgdata/geneseq/geneseq/NA2000.DAT.AAA80613 +			692.00	1401.22	
	/SID58/cgcgdata/geneseq/geneseq/NA2000.DAT.AAA57119 +			692.00	1400.87	
	/SID58/cgcgdata/geneseq/geneseq/NA1997.DAT.AAX78509 +			676.00	1368.99	
	/SID58/cgcgdata/geneseq/geneseq/NA2001.DAT.AAD04006 +			406.00	808.88	
	/SID58/cgcgdata/geneseq/geneseq/NA2001.DAT.AAD04004 +			406.00	808.88	
	/SID58/cgcgdata/geneseq/geneseq/NA2000.DAT.AAA57118 +			393.50	777.33	
	/SID58/cgcgdata/geneseq/geneseq/NA2000.DAT.AAA42549 +			347.00	699.44	
	/SID58/cgcgdata/geneseq/geneseq/NA2001.DAT.AAD04007 +			344.00	681.44	
	/SID58/cgcgdata/geneseq/geneseq/NA2001.DAT.AAD04005 +			344.00	681.28	
	/SID58/cgcgdata/geneseq/geneseq/NA1998.DAT.AAV19148 +			325.50	650.44	
	/SID58/cgcgdata/geneseq/geneseq/NA1998.DAT.AAV19149 +			325.50	647.77	
	/SID58/cgcgdata/geneseq/geneseq/NA1998.DAT.AAV19147 +			313.50	627.11	
	/SID58/cgcgdata/geneseq/geneseq/NA2000.DAT.AAH68150 +			302.50	596.33	
	/SID58/cgcgdata/geneseq/geneseq/NA2000.DAT.AAA15931 +			302.50	594.77	
	/SID58/cgcgdata/geneseq/geneseq/NA2000.DAT.AAA15917 +			302.50	594.44	
	/SID58/cgcgdata/geneseq/geneseq/NA2000.DAT.AAZ65250 +			302.00	594.33	
	/SID58/cgcgdata/geneseq/geneseq/NA2001.DAT.AAZ62743 +			298.50	582.11	
	/SID58/cgcgdata/geneseq/geneseq/NA2000.DAT.AAA80663 +			227.50	451.88	
	/SID58/cgcgdata/geneseq/geneseq/NA2001.DAT.AAH68153 +			181.00	339.77	
	/SID58/cgcgdata/geneseq/geneseq/NA2001.DAT.AAH68134 +			181.00	284.11	
	/SID58/cgcgdata/geneseq/geneseq/NA2000.DAT.AAA80662 +			172.00	343.00	
	/SID58/cgcgdata/geneseq/geneseq/NA2001.DAT.AAD02742 +			154.00	293.00	
	/SID58/cgcgdata/geneseq/geneseq/NA2001.DAT.AAI41251 +			148.00	287.55	
	/SID58/cgcgdata/geneseq/geneseq/NA2001.DAT.AAD04002 +			115.00	213.99	
	/SID58/cgcgdata/geneseq/geneseq/NA2001.DAT.AAH52178 +			114.00	221.77	
	/SID58/cgcgdata/geneseq/geneseq/NA2001.DAT.AAI34290 +			111.50	225.00	
	/SID58/cgcgdata/geneseq/geneseq/NA2001.DAT.AAD04003 +			86.00	164.33	
	/SID58/cgcgdata/geneseq/geneseq/NA1999.DAT.AAF33946 +			86.00	135.65	
	/SID58/cgcgdata/geneseq/geneseq/NA2001.DAT.AAF59633 +			86.00	135.55	
	/SID58/cgcgdata/geneseq/geneseq/NA2001.DAT.AAD10020 +			84.00	144.33	
	/SID58/cgcgdata/geneseq/geneseq/NA1995.DAT.AAQ93138 +			84.00	135.88	
	/SID58/cgcgdata/geneseq/geneseq/NA1995.DAT.AAQ93484 +			84.00	135.55	
	/SID58/cgcgdata/geneseq/geneseq/NA1997.DAT.AQT93873 +			84.00	135.55	
	/SID58/cgcgdata/geneseq/geneseq/NA2000.DAT.AAZ60181 +			84.00	135.55	

Georgiev G, Kiselev S, Ostermann E, Prokhorchouk E;
WPI: 1999-120887/10.
P-PSDB; AAY00770.
New nucleic acid encoding tag7 - used to inhibit tumour growth
induce apoptosis, for treatment of carcinoma, sarcoma, mel,
leukaemia
Claim 3; Fig 1; 138pp; English.
This sequence encodes the murine tag7 of the invention. CC
the tag7 DNA sequence are used to express recombinant tag7 CC
to produce and purify antibodies: to inhibit growth of man CC
especially for treating carcinoma (of liver, ovary, breast, CC
prostate, colon/rectum, bladder, testis, stomach, pancreas, CC
and neck, squamous cell carcinoma or teratocarcinoma), sarcoma CC
osteo- or fibro-sarcomas), melanoma or leukaemia; and as a CC
weight marker. The tag7 polypeptide inhibits tumour growth CC
apoptosis. The tag7 coding sequences are also useful as pri CC
mapping and detection of tag7 gene expression, and as pri CC
against tag7 are used as reagents for detecting tag7; as a CC
tag7; for isolating tag7 and therapeutically to inhibit or CC
metastasis.
Sequence 549 BP; 117 A; 161 C; 151 G; 120 T; 0 other;
alignment_scores:
Quality: 996.00 Length: 182
Ratio: 5.473 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-09-462-625-2 x AAX21819 ..
Align seg 1/1 to: AAX21819 from: 1 to: 549
1 MetLeuPheAlaCysAlaLeuLeuAlaLeuGlyLeuAlaThrSerCy 17
1 ATGTTCTTTTGCCCTGTGCTCTCCCTTCCTGGCTCTGGCAACTCTG 11

Wed Dec 19 08:45:26 2001

us-09-462-625-2.p2n.rng

```

17 sSerPheIleValProArgSerGluTyrArgAlaLeuProSerGluCysS 34
18 |
19 |
20 |
21 |
22 |
23 |
24 |
25 |
26 |
27 |
28 |
29 |
30 |
31 |
32 |
33 |
34 |
35 |
36 |
37 |
38 |
39 |
40 |
41 |
42 |
43 |
44 |
45 |
46 |
47 |
48 |
49 |
50 |
51 CAGTTTCATCGTGCCTCCGAGTGGAGGCGCTGCCATCCGAGTGT 100
52 |
53 |
54 |
55 |
56 |
57 |
58 |
59 |
60 |
61 |
62 |
63 |
64 |
65 |
66 |
67 |
68 |
69 |
70 |
71 |
72 |
73 |
74 |
75 |
76 |
77 |
78 |
79 |
80 |
81 |
82 |
83 |
84 |
85 |
86 |
87 |
88 |
89 |
90 |
91 |
92 |
93 |
94 |
95 |
96 |
97 |
98 |
99 |
100 |
101 CTAGCCGCTGGGCGACCCAGTACGTCACGAGGCGCCGCAATGT 150
102 |
103 |
104 |
105 |
106 |
107 |
108 |
109 |
110 |
111 |
112 |
113 |
114 |
115 |
116 |
117 |
118 |
119 |
120 |
121 |
122 |
123 |
124 |
125 |
126 |
127 |
128 |
129 |
130 |
131 |
132 |
133 |
134 |
135 |
136 |
137 |
138 |
139 |
140 |
141 |
142 |
143 |
144 |
145 |
146 |
147 |
148 |
149 |
150 |
151 GlySerPheCysAsnSerProAspSerCysGluGlnGlnAlaArgAsnVa 67
152 |
153 |
154 |
155 |
156 |
157 |
158 |
159 |
160 |
161 |
162 |
163 |
164 |
165 |
166 |
167 |
168 |
169 |
170 |
171 |
172 |
173 |
174 |
175 |
176 |
177 |
178 |
179 |
180 |
181 |
182 |
183 |
184 |
185 |
186 |
187 |
188 |
189 |
190 |
191 |
192 |
193 |
194 |
195 |
196 |
197 |
198 |
199 |
200 |
201 GGCAGCTTCTGCAACAGCCCGACTCTGTGAACAGAGGCGCCGCAATGT 200
202 |
203 |
204 |
205 |
206 |
207 |
208 |
209 |
210 |
211 |
212 |
213 |
214 |
215 |
216 |
217 |
218 |
219 |
220 |
221 |
222 |
223 |
224 |
225 |
226 |
227 |
228 |
229 |
230 |
231 |
232 |
233 |
234 |
235 |
236 |
237 |
238 |
239 |
240 |
241 |
242 |
243 |
244 |
245 |
246 |
247 |
248 |
249 |
250 |
251 |
252 |
253 |
254 |
255 |
256 |
257 |
258 |
259 |
260 |
261 |
262 |
263 |
264 |
265 |
266 |
267 |
268 |
269 |
270 |
271 |
272 |
273 |
274 |
275 |
276 |
277 |
278 |
279 |
280 |
281 |
282 |
283 |
284 |
285 |
286 |
287 |
288 |
289 |
290 |
291 |
292 |
293 |
294 |
295 |
296 |
297 |
298 |
299 |
300 |
301 |
302 |
303 |
304 |
305 |
306 |
307 |
308 |
309 |
310 |
311 |
312 |
313 |
314 |
315 |
316 |
317 |
318 |
319 |
320 |
321 |
322 |
323 |
324 |
325 |
326 |
327 |
328 |
329 |
330 |
331 |
332 |
333 |
334 |
335 |
336 |
337 |
338 |
339 |
340 |
341 |
342 |
343 |
344 |
345 |
346 |
347 |
348 |
349 |
350 |
351 |
352 |
353 |
354 |
355 |
356 |
357 |
358 |
359 |
360 |
361 |
362 |
363 |
364 |
365 |
366 |
367 |
368 |
369 |
370 |
371 |
372 |
373 |
374 |
375 |
376 |
377 |
378 |
379 |
380 |
381 |
382 |
383 |
384 |
385 |
386 |
387 |
388 |
389 |
390 |
391 |
392 |
393 |
394 |
395 |
396 |
397 |
398 |
399 |
400 |
401 |
402 |
403 |
404 |
405 |
406 |
407 |
408 |
409 |
410 |
411 |
412 |
413 |
414 |
415 |
416 |
417 |
418 |
419 |
420 |
421 |
422 |
423 |
424 |
425 |
426 |
427 |
428 |
429 |
430 |
431 |
432 |
433 |
434 |
435 |
436 |
437 |
438 |
439 |
440 |
441 |
442 |
443 |
444 |
445 |
446 |
447 |
448 |
449 |
450 |
451 |
452 |
453 |
454 |
455 |
456 |
457 |
458 |
459 |
460 |
461 |
462 |
463 |
464 |
465 |
466 |
467 |
468 |
469 |
470 |
471 |
472 |
473 |
474 |
475 |
476 |
477 |
478 |
479 |
480 |
481 |
482 |
483 |
484 |
485 |
486 |
487 |
488 |
489 |
490 |
491 |
492 |
493 |
494 |
495 |
496 |
497 |
498 |
499 |
500 |
501 |
502 |
503 |
504 |
505 |
506 |
507 |
508 |
509 |
510 |
511 |
512 |
513 |
514 |
515 |
516 |
517 |
518 |
519 |
520 |
521 |
522 |
523 |
524 |
525 |
526 |
527 |
528 |
529 |
530 |
531 |
532 |
533 |
534 |
535 |
536 |
537 |
538 |
539 |
540 |
541 |
542 |
543 |
544 |
545 |
546 |
547 |
548 |
549 |
550 |
551 |
552 |
553 |
554 |
555 |
556 |
557 |
558 |
559 |
560 |
561 |
562 |
563 |
564 |
565 |
566 |
567 |
568 |
569 |
570 |
571 |
572 |
573 |
574 |
575 |
576 |
577 |
578 |
579 |
580 |
581 |
582 |
583 |
584 |
585 |
586 |
587 |
588 |
589 |
590 |
591 |
592 |
593 |
594 |
595 |
596 |
597 |
598 |
599 |
600 |
601 |
602 |
603 |
604 |
605 |
606 |
607 |
608 |
609 |
610 |
611 |
612 |
613 |
614 |
615 |
616 |
617 |
618 |
619 |
620 |
621 |
622 |
623 |
624 |
625 |
626 |
627 |
628 |
629 |
630 |
631 |
632 |
633 |
634 |
635 |
636 |
637 |
638 |
639 |
640 |
641 |
642 |
643 |
644 |
645 |
646 |
647 |
648 |
649 |
650 |
651 |
652 |
653 |
654 |
655 |
656 |
657 |
658 |
659 |
660 |
661 |
662 |
663 |
664 |
665 |
666 |
667 |
668 |
669 |
670 |
671 |
672 |
673 |
674 |
675 |
676 |
677 |
678 |
679 |
680 |
681 |
682 |
683 |
684 |
685 |
686 |
687 |
688 |
689 |
690 |
691 |
692 |
693 |
694 |
695 |
696 |
697 |
698 |
699 |
700 |
701 |
702 |
703 |
704 |
705 |
706 |
707 |
708 |
709 |
710 |
711 |
712 |
713 |
714 |
715 |
716 |
717 |
718 |
719 |
720 |
721 |
722 |
723 |
724 |
725 |
726 |
727 |
728 |
729 |
730 |
731 |
732 |
733 |
734 |
735 |
736 |
737 |
738 |
739 |
740 |
741 |
742 |
743 |
744 |
745 |
746 |
747 |
748 |
749 |
750 |
751 |
752 |
753 |
754 |
755 |
756 |
757 |
758 |
759 |
760 |
761 |
762 |
763 |
764 |
765 |
766 |
767 |
768 |
769 |
770 |
771 |
772 |
773 |
774 |
775 |
776 |
777 |
778 |
779 |
780 |
781 |
782 |
783 |
784 |
785 |
786 |
787 |
788 |
789 |
790 |
791 |
792 |
793 |
794 |
795 |
796 |
797 |
798 |
799 |
800 |
801 |
802 |
803 |
804 |
805 |
806 |
807 |
808 |
809 |
810 |
811 |
812 |
813 |
814 |
815 |
816 |
817 |
818 |
819 |
820 |
821 |
822 |
823 |
824 |
825 |
826 |
827 |
828 |
829 |
830 |
831 |
832 |
833 |
834 |
835 |
836 |
837 |
838 |
839 |
840 |
841 |
842 |
843 |
844 |
845 |
846 |
847 |
848 |
849 |
850 |
851 |
852 |
853 |
854 |
855 |
856 |
857 |
858 |
859 |
860 |
861 |
862 |
863 |
864 |
865 |
866 |
867 |
868 |
869 |
870 |
871 |
872 |
873 |
874 |
875 |
876 |
877 |
878 |
879 |
880 |
881 |
882 |
883 |
884 |
885 |
886 |
887 |
888 |
889 |
890 |
891 |
892 |
893 |
894 |
895 |
896 |
897 |
898 |
899 |
900 |
901 |
902 |
903 |
904 |
905 |
906 |
907 |
908 |
909 |
910 |
911 |
912 |
913 |
914 |
915 |
916 |
917 |
918 |
919 |
920 |
921 |
922 |
923 |
924 |
925 |
926 |
927 |
928 |
929 |
930 |
931 |
932 |
933 |
934 |
935 |
936 |
937 |
938 |
939 |
940 |
941 |
942 |
943 |
944 |
945 |
946 |
947 |
948 |
949 |
950 |
951 |
952 |
953 |
954 |
955 |
956 |
957 |
958 |
959 |
960 |
961 |
962 |
963 |
964 |
965 |
966 |
967 |
968 |
969 |
970 |
971 |
972 |
973 |
974 |
975 |
976 |
977 |
978 |
979 |
980 |
981 |
982 |
983 |
984 |
985 |
986 |
987 |
988 |
989 |
990 |
991 |
992 |
993 |
994 |
995 |
996 |
997 |
998 |
999 |
1000 |

```

CC PRO polynucleotide and protein sequences given in the exemplification of
 CC the present invention.
 XX Sequence 697 BP; 136 A; 250 C; 184 G; 127 T; 0 other;
 SO

alignment_scores:
 Quality: 692.00 Length: 189
 Ratio: 4.325 Gaps: 1

Percent Similarity: 84.656 Percent Identity: 67.196

alignment_block:

US-09-462-625-2 x AAC58104 ..

Align seg 1/1 to: AAC58104 from: 1 to: 697

1 MetLeuPheAlaCysAlaLeuLeuAlaLeuGlyLeuAlaThrSer... 16
 41 ATGCTGCTTGGCTCTCCAGCCCTCTCGACTCGGAGCGGCTCA 90
 17CysSerPheIleValProArgSerGluT 26
 91 GGAGACAGAGACCCGGCTGCTGCAGCCCATAGTGCCCGGACGAGT 140
 26 rpArgAlaLeuProSerGluCysSerSerArgLeuGlyHisProValArg 42
 141 GGAAGCCCTGGCATCAGAGTGGCGGCAGCACCTGAGCCTGCCCTTACGC 190
 43 TyrValValIleSerHisThrAlaGlySerPheCysAsnSerProAspSe 59
 191 TATGTGTGTGTATCGCACGGCGGCAGCAGCTGCAACACCCCGCCCTC 240
 59 rCysGluGlnGlnAlaArgAsnValGlnHisTyrHisLysAsnGluLeuG 76
 241 GTCCAGCAGCAGCCCGGGAATGTGCAGCACTACCATGAAGACACTGG 290
 76 lyTrpCysAspValAlaTyrAsnPheLeuIleGlyGluAspGlyHisVal 92
 291 GCTGTGTCGACGTCGCTACACTTCTGATTGGAGAGACGGGCTCGTA 340
 93 TyrGluGlyArgGlyTrpAsnIleLysGlyAspHisThrGlyProIleTr 109
 341 TACGAGGGCCGTCGTGGAACTTCACGGGTGCCACTCAGTCAGTCATTATG 390
 109 pAsnProMetSerIleGlyIleThrPheMetGlyAsnPheMetAspArgV 126
 391 GAACCCCATGTCATTGGCATCGACTTCATGGGCACACTACATGGATCGG 440
 126 alProAlaLysArgAlaLeuArgAlaAlaLeuAsnLeuGluCysGly 142
 441 TGCCACACCCCGGCTCCGGGACGCCAGGGCTCTACTGGCCTCGGT 490
 143 ValSerArgGlyPheLeuArgSerAsnTyrGluValLysGlyHisArgAs 159
 491 GTGCTCAGGAGGCCCTGAGGTCCAACTATGTCTCAAGGACACCGGGA 540
 159 pValGlnSerThrLeuSerProGlyAspGlnLeuTyrGlnValIleGlnS 176
 541 TGTGACGGTACACTCTCTCCAGGCAACCACTCTACCACTCATCCAGA 590
 176 erTrpGluHisTyrArg 181
 591 ATTGGCCACACTACCCG 607

seq_name: /SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AAA37082

seq_documentation_block:

ID AAA37082 standard; cDNA; 697 BP.

XX

AC AAA37082;

XX

DT 08-AUG-2000 (first entry)

XX

117 rPheMetGlyAsnPheMetAspArgValProAlaLysArgAlaLeuArgA 134
 386 CTTATGGGGAACTTATGACCGGTA...CGCAAGAGCGCCCTCGTG 432
 134 laAlaLeuAsnLeuGluCysGlyValSerArgGlyPheLeuArgSer 150
 433 CTGCCCTAAATCTTCGGATCTGGGGTGTCTGGGGCTCTCTGAGATCC 482
 151 AsnTyrGluValLysGlyHisArgAspValGlnSerThrLeuSerProGl 167
 483 AACTATGAGTCAAGAGACACCGGATGTGCAAGCACTCTCTCTCCAGG 532
 167 yAspGlnLeuTyrGlnValIleGlnSerTrpGluHisTyrArgGlu 182
 533 TGACCAACTCTATCAGTCTATCAAGCTCGAAGCACTACCGAGAG 578

seq_name: /SID88/gcgdata/geneseq/geneseq/NA2000.DAT:AA58104

seq_documentation_block:

ID AAC58104 standard; cDNA; 697 BP.

XX

AC AAC58104;

XX

DT 25-JAN-2001 (first entry)

XX

DE Human PRO1269 nucleotide sequence SEQ ID NO:6.

XX

KW Human; tumour; diagnosis; neoplastic disease; proliferation; cancer;
 KW Identification; tumorigenesis; anticancer; detection; ss.

XX

OS Homo sapiens.

XX

PN WO200053750-A1.

XX

PD 14-SEP-2000.

XX

PF 02-DEC-1999; 99WO-US28551.

XX

PR 08-MAR-1999; 99WO-US05028.

XX

PR 01-SEP-1999; 99WO-US20111.

XX

PR 29-OCT-1999; 99US-0162506.

XX

PR 30-NOV-1999; 99WO-US28313.

XX

PR 01-DEC-1999; 99WO-US28634.

XX

PA (GETH) GENENTECH INC.

XX

PI Botstein D, Goddard A, Gurney AL, Roy MA, Watanabe CK, Wood WI;

XX

DR WPI; 2000-594320/56.

XX

DR P-PSDB; AAB24022.

XX

XX Antibodies specific for PRO polypeptides, used to diagnose and inhibit
 the growth of tumors in mammals, and to identify inhibitors of PRO
 polypeptide activity or expression -

XX

PS Claim 50; Fig 3; 226pp; English.

XX

CC The present invention describes an antibody that binds to a human
 protein (I) selected from: PRO381; PRO1269; PRO1410; PRO1755; PRO1780;
 PRO3434; PRO1927; PRO3567; PRO1295; PRO1293; PRO3444; PRO4354;
 PRO4397; PRO4507; PRO1555; PRO1096; PRO2038; and PRO2262. (I) has
 anticancer activity and can be used to diagnose tumours in mammals, by
 detecting complex formation when the antibody is contacted with test
 cells. Increased expression of genes encoding (I) can also be detected
 to diagnose tumours. Agents which inhibit the activity of (I),
 especially the antibodies, or an antisense oligonucleotide which
 hybridises to genes encoding (I), can be used to inhibit tumour growth,
 preferably by inducing cell death. Methods from the present invention
 can be used to identify compounds which inhibit the biological activity
 of (I). AAC58019 to AAC58102 represent PCR primers and hybridisation
 probes used in examples from the present invention for human PRO
 sequences. AAC58103 to AAC58122 and AAB24021 to AAB24040 represent human


```

alignment_scores:
  Quality: 692.00      Length: 189
  Ratio: 4.325        Gaps: 1
  Percent Similarity: 84.656  Percent Identity: 67.196

alignment_block:
  US-09-462-625-2 x AAX21820  ..

  Align seg 1/1 to: AAX21820 from: 1 to: 718

1 MetLeuPheAlaCysAlaLeuLeuAlaLeuLeuGlyLeuAlaThrSer.. 16
|||||:|||||  |||||:|||||  |||||:|||||  |||||:|||||
68 ATGCTGCTGGCTGGCTCTCCCGACGCTCTTCGACTCGAGCGGGTCA 117
|||||:|||||  |||||:|||||  |||||:|||||  |||||:|||||
17 .....CysSerPheIleValProArgSerGluT 26
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
118 GGAGCAGACAGACCCGGCGCTGTCGAGCGCCCATAGTGCCTCGGAACGAGT 167
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
26 pArgAlaLeuProSerGluCysSerSerArgLeuGlyHisProValArg 42
|||||:|||||  |||||:|||||  |||||:|||||  |||||:|||||
168 GGAAGCGCTGCATCAGAGTGGCCGACGACCTTGAGCCCTGCCCTTACGC 217
|||||:|||||  |||||:|||||  |||||:|||||  |||||:|||||
43 TyrValValIleSerHisThrAlaGlySerPheCysAsnSerProAspSe 59
|||||:|||||  |||||:|||||  |||||:|||||  |||||:|||||
218 TATGTGGTGGTATCGCACACGGCGGCAGCAGCTGCAACACCCCGCGCTC 267
|||||:|||||  |||||:|||||  |||||:|||||  |||||:|||||
59 rCysGluGlnGlnAlaArgAsnValGlnHisTyrHisLysAsnGluLeuG 76
|||||:|||||  |||||:|||||  |||||:|||||  |||||:|||||
268 GTGCCAGCAGCAGCGCCGGAATGTGCAGCACTACCACATGAAGACACTGG 317
|||||:|||||  |||||:|||||  |||||:|||||  |||||:|||||
76 lYTrpCysAspValAlaTyrAsnPheLeuIleGlyGluAspGlyHisVal 92
|||||:|||||  |||||:|||||  |||||:|||||  |||||:|||||
318 GTGTGGTGGCAGTGGGTACACTCTCTGATTGAGAAGACGGGCTCGTA 367
|||||:|||||  |||||:|||||  |||||:|||||  |||||:|||||
93 TyrGluGlyArgGlyTrpAsnIleLysGlyAspHisThrGlyProIleTr 109
|||||:|||||  |||||:|||||  |||||:|||||  |||||:|||||
368 TACGAGGCGGTGGCTGGAACCTTCCACGGGTGCCCACTCAGGTCACTTATG 417
|||||:|||||  |||||:|||||  |||||:|||||  |||||:|||||
109 pAsnProMetSerIleGlyIleThrPheMetGlyAsnPheMetAspArgV 126
|||||:|||||  |||||:|||||  |||||:|||||  |||||:|||||
418 GAACCCCATGTCATTGGCATCAGCTTCATGGGCAACTACATGGATCGGG 467
|||||:|||||  |||||:|||||  |||||:|||||  |||||:|||||
126 aProAlaLysArgAlaLeuArgAlaAlaLeuAsnLeuLeuGluCysGly 142
|||||:|||||  |||||:|||||  |||||:|||||  |||||:|||||
468 TGCCCCACACCCCGACCATCCGGCGCAGCCCGAGGGTCTACTGGCCTCGGCT 517
|||||:|||||  |||||:|||||  |||||:|||||  |||||:|||||
143 ValSerArgGlyPheLeuArgSerAsnTyrGluValLysGlyHisArgAs 159

```

```

|||||:||||| |||||||:|||||:|||||:|||||:|||||
518 GTGGCTCAGGAGCCCTGAGGTCCAACTATGTCTCAAGGACACCGGA 567
159 pValGlnSerThrLeuSerProGlyAspGlnLeuTyrGlnValIleGlns 176
||||| |||||||:|||||:|||||:|||||:|||||:|||||:|||||
568 TGTGCAGCGTACACTCTCTCCAGGCAACCACTCTCCACCTCATCCAGA 617
176 erTrrpGluHisTyrArg 181
:||||| |||||||
618 ATTGGCCACACTACCGC 634

```

seq_name: /SDS8/gcgdata/geneseq/geneseq/NA2000.DAT:AAA80613

seq_documentation_block:

ID AAA80613 standard; cDNA; 726 BP.

XX

AC AAA80613;

XX

DT 21-NOV-2000 (first entry)

XX

DE Human Htag7 secreted protein gene #8.

XX

KW Secreted protein; immunosuppressant; anti-inflammatory; antiarthritic;
 KW antirheumatic, dermatological; antiproliferative; antiarteriosclerotic;
 KW anticancer; vulnary; antiviral; antibacterial; antifungal;
 KW immune disorder; Addison's disease; rheumatoid arthritis; dermatitis;
 KW multiple sclerosis; inflammatory disorder; inflammatory bowel disease;
 KW Crohn's disease; nephritis; hyperproliferative disorder;
 KW cardiovascular disorder; coronary arteriosclerosis; myocarditis; cancer;
 KW melanoma; lymphoma; wound healing; human; ss.

XX

OS Homo sapiens.

XX

PN WO200029435-A1.

XX

PD 25-MAY-2000.

XX

XX 27-OCT-1999; 99WO-US25031.

XX

PR 28-OCT-1998; 98US-0105971.

XX

PA (HUMA-) HUMAN GENOME SCI INC.

XX

PI Ni J, Ruben SM, Olsen HS, Young PE, Kenny JJ, Moore PA, Wei Y;

PI Greene JM;

XX

XX WPI; 2000-387742/33.

DR

DR P-PSDB; AAB25583.

XX

PT Isolated nucleic acid molecules encoding human secreted proteins are
 PT used for the prevention, amelioration and treatment of autoimmune,
 PT inflammatory, hyperproliferative and cardiovascular disorders, cancer,
 PT wounds, and infectious diseases.

XX

PS Claim 1: Figure 34; 803pp; English.

XX

CC The present invention relates to 12 secreted human proteins and the
 CC nucleotide sequences encoding them. The polynucleotide sequences given
 CC in AAA80606-A80623 encode the 12 secreted protein sequences given in
 CC AAB25576-B25593. The human secreted proteins have various activities
 CC dependent on the tissues in which they are expressed. Examples of the
 CC activities of the proteins include: immunosuppressant;
 CC anti-inflammatory; antirheumatic; antirheumatic, dermatological;
 CC antiproliferative; antiarthritic; antirheumatic; anticancer; vulnary;
 CC antiviral; antibacterial; and antifungal activity. The proteins,
 CC polypeptides, agonists and antagonists may be used to treat prevent
 CC and/or diagnose various disease, disorders and conditions examples of
 CC which include: immune disorders e.g. Addison's disease, rheumatoid
 CC arthritis, dermatitis, and multiple sclerosis; inflammatory disorders
 CC e.g. inflammatory bowel disease, Crohn's disease and nephritis;
 CC hyperproliferative disorders such as paraneoplasias and purpura;
 CC cardiovascular disorders e.g. coronary arteriosclerosis and myocarditis;
 CC cancer e.g. melanoma and lymphoma. The proteins and polynucleotide

CC sequences may also be used in wound healing and the treatment of
 CC infectious diseases. The human secreted protein gene #8 and protein
 CC sequences are represented in sequences AAA80613 and AAB25583. Sequences
 CC AAA80662-A80663 represent genes related to the secreted protein gene#8.
 XX
 SQ Sequence 726 BP; 161 A; 251 C; 188 G; 126 T; 0 other;

alignment_scores:
 Quality: 692.00 Length: 189
 Ratio: 4.325 Gaps: 1
 Percent Similarity: 84.656 Percent Identity: 67.196

alignment_block:

US-09-462-625-2 x AAA80613

Align seg 1/1 to: AAA80613 from: 1 to: 726

```

1 MetLeuPheAlaCysAlaLeuLeuAlaLeuLeuGlyLeuAlaThrSer.. 16
47 ATGCTGCTTGGCTGGGCTCTCCAGCAGCCTCTTCGACTCGGAGCGGCTCA 96
17 .....CysSerPheIleValProArgSerGluT 26
97 GGAGACAGAAAGACCGCGCTGCTGCAGCCCCCATAGTCCCGGAAACGAGT 146
26 rpArgAlaLeuProSerGluCysSerArgSerArgLeuGlyHisProValArg 42
147 GGAAGGCCCTGGCATCAGATGCGCCAGCACCTGAGCCTGCCCTTACGC 196
43 TyrValValIleSerHisThrAlaGlySerPheCysAsnSerProAspSe 59
197 TATGTGTGTGTATCGCACAGCGGCGGAGCAGTGTCAACACCCCGGCTC 246
59 rCysGluGlnGlnAlaArgAsnValGlnHisTyrHisLysAsnGluLeuG 76
247 GTGCCAGCAGCAGCGCGGGAATGTGCAGCAGCTACCATCATGAAGACACTGG 296
76 lyTrpCysAspValAlaTyrAsnPheLeuIleGlyGluAspGlyHisVal 92
297 GCTGTGGCGAGCTGGGTACAACTTCTGATTGCGAGAAGACGGGCTCGTA 346
93 TyrGluGlyArgGlyTyrAsnIleGlyAspHisThrGlyProIleTr 109
347 TACGAGGCGCGTGGCTGGAACCTTCACGGGTGCCCACTCAGCTCATTATG 396
109 pAsnProMetSerIleGlyIleThrPheMetGlyAsnPheMetAspArgV 126
397 GAACCCCATGTCCATTGGCATCAGCTTCTGAGGCACTGATGATGATGATG 446
126 alProAlaLysArgAlaLeuArgAlaLeuAsnLeuLeuGluCysGly 142
447 TGCCACACCCCGGCGCATCCGCGGCGAGCCAGGCTCTACTGGCTCGGCT 496
143 ValSerArgGlyPheLeuArgSerAsnTyrGluValLysGlyHisArgAs 159
497 GTGCTCAGGAGCGGCTGAGGTCCAACTATGTCTCAAGGACACCGGGA 546
159 pValGlnSerThrLeuSerProGlyAspGlnLeuTyrGlnValIleGlns 176
547 TGTGAGGGTACACTCTCTCCAGGCAACCACTCTACCATCATCCAGA 596
176 erTrrpGluHisTyrArg 181
597 ATTGGCCACACTACCGC 613

```

seq_name: /SDS8/gcgdata/geneseq/geneseq/NA2000.DAT:AAA51719

seq_documentation_block:

ID AAA51719 standard; cDNA; 749 BP.

XX

AC AAA51719;

XX

Wed Dec 19 08:45:26 2001

us-09-462-625-2.p2n.rng

31-OCT-2000 (first entry)
Chondrosarcoma peptidoglycan recognition protein-like protein cDNA.
Peptidoglycan recognition protein-like protein; PGRP; PGRP-C; regulator;
chondrosarcoma; immunosuppressive; antibacterial; anti-apoptotic;
tumour necrosis factor; TNF; anti-infectious; mononuclear phagocyte;
inhibitor; protein co-ordinate data; ss.

Homo sapiens.
Key Location/Qualifiers
CDS 55..645
FT /*tag= a
FT /product= PGRP-like_protein
FT sig_peptide 55..117
FT /*tag= b
FT mat_peptide 118..642
FT /*tag= c
WO200039327-A1.
06-JUL-2000.
22-DEC-1999; 99WO-US30736.
23-DEC-1998; 98US-0113809.
(HUMA-) HUMAN GENOME SCI INC.
Rosen CA, Ruben SM, Young PE, Olsen HS;
WPI; 2000-452414/39.
P-PSDB; AAY96964.
Polynucleotide encoding peptidoglycan recognition protein-like protein,
antibodies specific to it useful for preventing, treating conditions
e.g. endotoxemic shock and auto-immune disorders and infections in mammal
Claim 1; Fig 3; 191pp; English.

Novel human peptidoglycan recognition protein-like proteins (PGRP)
expressed by keratinocytes (PGRP-K), wound healing tissues (PGRP-W)
or chondrosarcoma tissue (PGRP-C) are disclosed. The PGRP DNA and
proteins are useful for preventing, treating or ameliorating a medical
condition in a mammal (claimed). PGRP is useful in augmenting the immune
system in such areas as immune recognition, antigen presentation and
immune system activation. Antibodies or antagonists directed against
these proteins may be useful in reducing or eliminating disorders
associated with tumour necrosis factor (TNF) and TNF-like cytokines,
such as endotoxemic shock and autoimmune disorders and for treating
infectious diseases including silicosis, sarcoidosis and idiopathic
pulmonary fibrosis.

Sequence 749 BP; 165 A; 259 C; 195 G; 130 T; 0 other;

alignment_scores:
Quality: 692.00 Length: 189
Ratio: 4.325 Gaps: 1
Percent Similarity: 84.656 Percent Identity: 67.196

alignment_block:
US-09-462-625-2 x AAA51719 ..
Align seg 1/1 to: AAA51719 from: 1 to: 749
1 MetLeuPheAlaCysAlaLeuLeuAlaLeuLeuGlyLeuAlaThrSer... 16
70 ATGCTGCTTGGCTCTCCAGGCTCTCTGACTCGGAGCGGCTCA 119
17CysSerPheIleValProArgSerGluT 26
||||| |||||||

120 GGAGACAGAGAGACCCGGCTGCTGCAGCCCATAGTGTCCCGGAGACGAGT 169
26 rpArgAlaLeuProSerGluCysSerSerArgLeuGlyHisProValArg 42
170 GGAAGGCCCTGGCATCAGATGGCGCCAGCACCTGAGCCTGCCCTTACGC 219
43 TyrValValIleSerHisThrAlaGlySerPheCysAsnSerProAspSe 59
220 TATGTGTGTATGCGACACGGCGGCGAGCTGCAACACCCCGGCTC 269
59 rCysGluGlnGlnAlaArgAsnValGlnHisTyrHisLysAsnGluLeuG 76
270 GTGCCAGCAGCAGCGCCGGAATGTGCAGCACTACCATGAGGACACATGG 319
76 lyTrpCysAspValAlaTyrAsnPheLeuIleGlyGluAspGlyHisVal 92
320 GCTGTGGGACGTGGGCTGACAACTTCTGATTGGAGAGACGGGCTCGTA 369
93 TyrGluGlyArgGlyTyrPAsnIleLysGlyAspHisThrGlyProIleTr 109
370 TACGAGGCGCGTGGCTGGAACCTTACGGGTGCCCACTCAGGTCACTATG 419
109 PAsnProMetSerIleGlyIleThrPheMetGlyAsnPheMetAspArgV 126
420 GAACCCCATGTCCATTGGCATCAGCTTTCATGGCACTACATGGATCGGG 469
126 alProAlaLysArgAlaLeuArgAlaAlaLeuAsnLeuLeuGluCysGly 142
470 TGCCACACACCCAGCCATCCGGCGAGCCAGGGTCTACTGGCTCGGCT 519
143 ValSerArgGlyPheLeuArgSerAsnTyrGluValLysGlyHisArgAs 159
520 GTGGCTCAGGGAGCCCTGAGGTCCAACCTATGTCTCAAGAGACACCGGA 569
159 pValGlnSerThrLeuSerProGlyAspGlnLeuTyrGlnValIleGlnS 176
570 TGTGACGCTTACACTCTCTCCAGGCAACCACTTACACCTACCTATCCAGA 619
176 erTrpGluHisTyrArg 181
620 ATGGGCCACACTACCGC 636
seq_name: /SID88/gcgdata/geneseq/geneseq/NAL1997.DAT: AAT78509
seq_documentation_block:
ID AAT78509 standard; cDNA; 688 BP.
XX
AC AAT78509;
XX
DT 18-FEB-1998 (first entry)
XX
DE Bovine granulocyte peptide A precursor cDNA.
XX
KW Antimicrobial peptide; antibiotic; antibacterial; antifungal;
KW fungicide; antiprotosoa; protozoacide; antiviral; virucide;
KW bovine granulocyte peptide A; Bgp-A; preservative; sepsis;
KW endotoxaemia; cattle; ss.
XX
OS Bos taurus.
XX
FH Key Location/Qualifiers
FT CDS 29..601
FT /*tag= a
FT sig_peptide 29..91
FT /*tag= b
FT mat_peptide 560..598
FT /*tag= c
FT /product= BGP-A
XX
XX WO9729765-A1.
XX
XX 21-AUG-1997.
XX


```

|||||::: |||||:::|||||:::|||||:::|||||
1048 ACCTTGTCCTCGGCAGGCTTGTATACATCATCAGCACCTGGCCTCA 1097
      .
179 sTVrArg 181
      . |:::
1098 TTTCAA 1104

seq_name: /SIDS8/gcndata/geneseq/geneseqn/NA2001.DAT:AAD04004

seq_documentation_block:
ID   AAD04004 standard; cDNA; 1128 BP.
XX   AC   AAD04004;
XX   AC   XX
XX   DT   XX
XX   DE   02-JUL-2001 (first entry)
XX   DE   Human full length granulocyte peptide homolog Zgpai cDNA #1
XX   DE   Humán; granulocyte peptide A; GP-A; Zgpai; cytostatic; anti-
XX   KW   vulnary; dermatological; anti-microbial; gastrointestinal;
XX   KW   pulmonary; dental carries; periodontal disease; gene therapy
XX   KW   Acquired Immune Deficiency Syndrome; epithelial wound; cyst
XX   KW   infection; urinary tract; respiratory; vaginal; lung; skin
XX   KW   exarian; rectal; chromosome 1; ss.

```

XX	Key	Location/Qualifiers
PH	CDS	1..1128
FT		/tag= a
FT		/product= "Human full length zgal protein #1 with linker sequence"
FT		
FT		

FT	linker sequence*
FT	1..51
FT	sig_peptide
FT	/*tag= b
FT	52..1125
FT	mat_peptide
FT	/*tag= c

	mat_peptide	52..1125	/tag=	/product=	"Mature human full length zgal protein #1"
FT					
FT					
FT					
XX					
PN					

FT	/product="Mature human full length zgpai protein"
XX	
XX	WO200129224-A2.
PN	
XX	
XX	26-APP-2001
XX	

26-APR-2001.	26-APR-2001.
20-OCT-2000; 2000WO-US29177.	20-OCT-2000; 2000WO-US29177.

XX	20-OCT-2000; 2000WO-US29177.
PF	
XX	
XX	20-OCT-1999; 99US-0160712.
PR	12-JUL-2000; 2000US-0218070.
PR	

XX
PR 20-OCT-1999; 99US-0160712.
PR 12-JUL-2000; 2000US-0218070.
XX
XX
FA (ZYMO) ZYMOGENETICS INC.

PR 12-JUL-2000; 200003-0218070.
XX
XX (ZYMO) ZYMOGENETICS INC.
PA
XX
PI Conklin DC, Adler DA, Fox BA;
XX

PA (ZYMO) ZYMOGENETICS INC.
XX
XX
PI Conklin DC, Adler DA, Fox BA;
XX
XX WPI: 2001-290918/30.
DR

PI Conklin DC, Adler DA, Fox BA;
XX
XX WPI: 2001-290918/30.
DR P-ESDB: AAE00692.
DR
XX
XX

DR WPI; 2001-290918/30.
DR P-PSDB; AAE00692.
XX New granulocyte peptide homolog, zppal polypeptide, for research applications, diagnosis and treatment of cancer, periodontal, PT
PT

XX New granulocyte peptide homolog, zgpal polypeptide, for research
PT applications, diagnosis and treatment of cancer, periodontal,
PT gastrointestinal disease, urinary tract, skin and lung infections
YV

XX applications, diagnosis and treatment of cancer, periodontal,
XX gastrointestinal disease, urinary tract, skin and lung infections
XX Claim 2; Page 102-105; 114pp; English.

XX Claim 2; Page 102-105; 114pp; English.
PS
XX
XX
CC The present sequence is a cDNA encoding human full length granulocyte
CC *α*-defensin protein. Zapal gene is located on human
CC chromosome 8p21.1 homologous to the human
CC *α*-defensin gene.

The present sequence is a cDNA encoding human full length granulocyte peptide (gp-A) homolog, Zgal protein. Zgal gene is located on human chromosome 1. Zgal polypeptides are useful for producing antibodies which are useful for detecting cancer. Zgal polypeptides having

peptide (GP-A) homolog, Zgpal protein. Zgpal gene is located on human chromosome 1. Zgpal polypeptides are useful for producing antibodies which are useful for detecting cancer. Zgpal polypeptides having anti-microbial activity are useful for treating dental caries, gastrointestinal disease, and urinary tract infection.

which are useful for detecting cancer. Zppal polypeptides having anti-microbial activity are useful for treating dental carries, periodontal disease, thrush, gastrointestinal disease, urinary tract infections, respiratory infections, vaginal infections, acquired immunodeficiency syndrome, and lung infections associated with cystic fibrosis.

periodontal disease, thrush, gastrointestinal disease, urinary tract infections, respiratory infections, vaginal infections, acquired immunodeficiency syndrome (AIDS) and lung infections associated with cystitis. Infection in skin and other epithelial wounds can be fatal. In vivo killing of *C. albicans* is successful for each type of infection.

infections, respiratory distress syndrome, and deficiency syndrome (AIDS) and lung infections associated with cystic fibrosis and prevention of infection in skin and other epithelial wounds. Zyxal-cytokine fusion proteins are useful for enhancing *in vivo* killing of target tissues (epithelial cancers, and more specifically lung, colorectal, and pancreatic cancers), and more specifically, fusion proteins that enhance the killing of tumor cells by cytotoxic T lymphocytes.

CC fibrosis and prevention of metastases.
CC zgpai-cytokine fusion proteins are useful for enhancing *in vivo* killing of target tissues (epithelial cancers, and more specifically lung, ovarian and rectal cancers). zgpai polypeptides, fragments, fusion proteins or agonists are useful in *in vitro* studies of exogenous

CC of target tissues (epithelial cancers; and bone metastases); 2ggal polypeptides, fragments, fusion
CC ovarian and rectal cancers). 2ggal polypeptides are useful in *in vitro* studies of exogenous
CC proteins or agonists

CC proteins or agonists are useful in *in vitro* studies of endogenous

microorganism infections such as bacterial, viral or fungal infection and also to study epithelial cell defense induction in cell culture. Zgalp antibodies, polynucleotides and polypeptides are useful for detection of zgalp polypeptide, mRNA or anti-zgalp antibodies, thus serving as markers for detecting genetic diseases or cancers. Zgalp sequences are useful as diagnostics in forensic DNA profiling and as probes or primers to clone 5' non-coding regions of a zgalp gene. Zgalp antibodies are useful for tagging cells that express zgalp, for screening expression libraries and as neutralizing antibodies or as antagonists to block zgalp activity in vitro and in vivo. Zgalp gene is also useful in gene therapy.

```
alignment_scores:      Length: 169
                        Quality: 406.00
                        Ratio: 3.355
                        Gaps: 2
                        Percent Similarity: 71.598
                        Percent Identity: 43.787
```

alignment_block:

US-09-462-625-2 X AAD04004

Align seq 1/1 to: AAD04004 from: 1 to: 1128

[illegible]

seq name: /STPS8/acgdata/geneseq/geneseq/NA2000.DAT:AAA51718

seq_documentation_block:
ID AAA51718 standard: cDNA: 1876 BP.

XX
ID AAK51/18 SLANGUAG, CDNR, 1970 DE

AC	AAA51718;	
XX	31-OCT-2000 (first entry)	
XX		
DE	Wound healing tissue peptidoglycan recognition protein-like protein CDNA.	
DE		
KW	Peptidoglycan recognition protein-like protein; PGRP; PGRP-W; regulator;	
KW	wound healing tissue; immunosuppressive; antibacterial; anti-apoptotic;	
KW	tumour necrosis factor; TNF; anti-infectious; mononuclear phagocyte;	
KW	inhibitor; protein co-ordinate data; ss.	
XX		
OS	Homo sapiens.	
XX		
XX	Key	Location/Qualifiers
PH	CDS	107..1213
FT		/*tag= a
FT		/product= PGRP-W
FT	sig_peptide	107..137
FT		/*tag= b
FT	mat_peptide	158..1210
FT		/*tag= c
XX		
XX	WO200039327-A1.	
PN		
XX	06-JUL-2000.	
XX		
PD		
XX	22-DEC-1999; 99WO-US30736.	
XX		
PF		
XX	23-DEC-1998; 98US-0113809.	
XX		
PR		
XX		
XX	(HUMA-) HUMAN GENOME SCI INC.	
XX		
PA		
XX		
PI	Rosen CA, Ruben SM, Young PE, Olsen HS;	
XX		
XX		
DR	WPI; 2000-452414/39.	
DR	P-PSDB; AAY96963.	
XX		
PT	Polynucleotide encoding peptidoglycan recognition protein-like protein,	
XX	antibodies specific to it useful for preventing, treating conditions	
PT	e.g. endotoxic shock and auto-immune disorders and infections in mammal	
PT		
PS	Claim 1. Fig 2A-B: 191bp; English.	
XX		

claim 1: Fig 2A-B; 191pp; English.

Novel human peptidoglycan recognition protein-like proteins (PGRP) expressed by keratinocytes (PGRP-K), wound healing tissues (PGRP-W) or chondrosarcoma tissue (PGRP-C) are disclosed. The PGRP DNA and proteins are useful for preventing, treating or ameliorating a medical condition in a mammal (claimed). PGRP is useful in augmenting the immune system in such areas as immune recognition, antigen presentation and immune system activation. Antibodies or antagonists directed against these proteins may be useful in reducing or eliminating disorders associated with tumour necrosis factor (TNF) and TNF-like cytokines, such as endotoxic shock and autoimmune disorders and for treating infectious diseases including silicosis, sarcoidosis and idiopathic pulmonary fibrosis.

XX
XX
SO

alignment_scores:		
Quality:	393.50	Length: 167
Ratio:	3.279	Gaps: 3
Percent Similarity:	71.856	Percent Identity: 44.910

alignment block:

US-09-462-625-2 x AAA51718

align seg 1/1 to: AAA51718 from: 1 to: 1876

[\[mailto:aleu@prose.org\]](#) 32

718 GCTTGGCCCCGGCATTTGTCCACGGTCTGTGTGGGAGCC.....AGGGA 761

```

32 ucYSer.SerArgLeuGlyHisProValArgTyrValValIleSerHis 48  
| : :: ||||:: | : ||::||:: ||::||:: ||::||  
762 GACCACGTGCCAGATCACTCCGCAGCGAAGTATGGCATTTATCCAC 811  
  
49 ThrAlaGlySerPheCysAsnSerProAspSerCysGluInGlnAlaAr 65  
||||||| |::||:: ||::||:: ||::||:: ||:  
812 ACTCGCGGAGGACTGCACAATTTCGTGATGAGTGCGCCCTGCTGGTCCG 861  
  
65 qasnValGlnHisTyrHisLysAsnGluLeuGlyTrpCysaspValalaT 82  
|||::||| :::: |::||| |::||| |::|||  
862 GGACATCCAGCTTTCATACATAGACAGCTCAAAGTCATCGCACATTGGTT 911  
  
82 yrAsnPheLeulleGlyCluaspgLyHisValTyrGluGlyArgGlyTrp 98  
||||||| |::||:: ||::||:: ||::||:: ||:  
912 ATAACCTTCCTGGTGGCCAGGATGCGGCATTTATGAAGGGGTGGGCTGG 961  
  
99 AsnileLysGlyASPHisThrGlyProIleTrpAsnProMetSerIleGl 115  
|||::||| :::: |::||| |::||| |::|||  
962 AATGTCCAAGCTCCTCCACC...CTGGCTACGATGACATATGCCCTGGG 1008  
  
115 ylleThrPheKetGlyAsnPheMetasPArgValProalalysArgAlaL 132  
||||||| |::||:: ||::||:: ||::||:: ||:  
1009 CATTAACCTTCATGGCACCCTTCACAGGTATACCCACCAATACTGCAGCAC 1058  
  
132 euAtqAlaAlaLeuAsnLeuLeuGluCysGlyValserArgglyPheLeu 148  
||||||| |::||:: ||::||:: ||::||:: ||:  
1059 TAGAGGCAGCCCCAACCTGTACCTGTCAGTGTGCCATGGTCAAGGGTACCTG 1108  
  
149 ArgSerAsnTyrrGluVallysGlyHisArgASPvalGlnSerThrlEuSe 165  
||||||| |::||:: ||::||:: ||::||:: ||:  
1109 ACTCCCAACTACTGCTGTGTGGGCCACAGTAGTGTCGCCGCAACCTTGTc 1158  
  
165 rPrOGlyaspGlnLeuTyrrGluValilleGlnSerTyrrGluHisTyraRg 181  
||||||| |::||:: ||::||:: ||::||:: ||:  
1159 TCTGTGGCAGGCTTGTACATCATCATCAGCACCCTGGCCTCATTTCAA 1207  
  
seq_name: /SIDBS/gcgdata/geneseq/geneseqn/NA2000.DAT.AAZ42549  
seq_documentation_block:  
Seq_AA242549 standard; cDNA: 380 BP.  
  
XX AC AAZ42549;  
XX EDT  
XX DE  
XX KW  
XX PPF  
XX PR  
XX RR  
XX RW  
XX WW  
XX YY  
XX ZZZ  
OS Homo sapiens.  
PN W09953051-A2.  
PPD 21-OCT-1999.  
PF 09-APR-1999; 99WO-IB00712.  
XX 09-APR-1998; 98US-0057719.  
PR 28-APR-1998; 98US-0069047.  
RR  
RW  
WW  
YY  
ZZ  
PA (GEST ) GENSET.  
PI Dumas Milne Edwards J., Duclert A, Giordano J;  
PX WPI: 2000-038446/03.  
XX P-PSDB; AAY64935.  
XX Novel secreted protein 5' expressed sequence tag sequences used in  
PT diagnostic, forensic, gene therapy, and chromosome mapping procedures
```


XX
sequence 1125 BP: 191 A: 117 C: 204 G: 146 T: 467 other;

alignment_scores:		
Quality:	344.00	Length: 166
Ratio:	3.215	Gaps: 2
Percent Similarity:	64.458	Percent Identity: 36.747

alignment block:

```
alignment_block:
  US-09-462-625-2 x AAD04005
```

Align seg 1/1 to: AAD04005 from: 1 to: 1125

16 SerCysSerPheIleValProArgSerGluTrpArgAlaLeuProSerGI 32
:::|||||:::|||||||::: ||| |
631 GCNTGYCCNGCGTGTCTCCNMGNWSNGTWTGGGNCNMGARGACNCA 680

32 uCysSerArgLeuGlyHisProValArgTyrValValIleSerHisT 49
:||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
681 YTGVCNNMGN..ATGCAGTTCNCGNCNAARTAYGGAATHATHCAAYA 727

49 hrAlaGlySerPheCysAsnSerProAspSerpCysGluGlnGlnAlaarg 65
|||||||SerPhe::: ||||| ||: ||: ||: ||: ||: ||: ||: ||:
728 CMCNGCNMGNANCNTGYAAAYATHWSNGAYGARTGYMGNTNYTINGNMGN 777

66 AsnValGlnHisTyrHisLysAsnGluLeuGlyTrpCypAspValalaTy 82
::::::::::: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
778 GAVATHCARMSNTTYTAYATHGAYGMNNTYNAARWSNTGYGAYATHGNGTA 827

82 rAsnPhelIleGlyGluAspGlyHisValTyrGluGlyArgGlyTrpA 99
||||||| ::::::::::| ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
828 YAAITYTYTNGNCARGAYGGNCNAHTHAYARGCGNTGNGTTGA 877

99 snIleLysGlyAspHisThrGlyProIleTrpAsnPrometSerIleGly 115
::: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
878 AYGTNCARGGNWSNSNACN...CCNGGNTAYGAYATHGCGNYTMGN 924

116 IleThrPheMetGlyAsnPhMetaspArgValProAlalysargAlale 132
::: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
925 ATHACNTTYATGGNACNTTYACNGGNATHCCNCCNCAAYGCGNCGNCGNYT 974

```

QM of: US-09-462-625-2 to: EST:*   out_format : pfs

Date: Dec 17, 2001 10:56 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:
-MODEL=frame+p2n.model      -DEV=xlp
-Q/cgn2.1\USPTO_spool\US09462625\runat_17122001_074947_26100/app_query.fasta_1.241
-DB=EST    -QFMT=fastap    -SUFFIX=p2n.lst    -GAPOP=12.000    -GAPEXT=4.000
-MINMATCH=0.100    -LOOPCL=0.000    -LOOPEXT=0.000    -CGAPOP=6.000
-CGAPEXT=0.050    -XGAPOP=10.000    -XGAPEXT=0.500    -FGAPOP=6.000
-FGAPEXT=7.000    -YGAPOP=10.000    -YGAPEXT=0.500    -DELOB=6.000
-DELETE=-7.000    -START=1    -MATRIX=biosum62    -TRANS-human40.cdi
-LIST=45    -MODALIGN=200    -THR_SCORE=pct    -THR_MAX=100    -THR_MIN=0
-ALIGN=15    -MODE=LOCAL    -OUTFMT=pfs    -NORM=ext    -MINLEN=0
-MAXLEN=2000000000    -USER=US09462625    -CGCN1_1_0    -NCPU=6    -ICPU=3
-LOGLOG    -NO_XLPXY    -WAIT    -THREADS=1

Search information block:
Query: US-09-462-625-2
Query length: 182
Database: EST*
Database sequences: 11351937
Database length: 1077921985
Search time (sec): 1548.670000

score_list:
Strd Orig zScore BScore Len Documentation ...
gb_est2:BGI174272 + 996.00 1848.18 8.8e-94 ! BGI174272 602334571P1 NCI_CGAP_M
gb_est2:BK1149595 + 996.00 1847.37 9.8e-94 ! BK1149595 602848402P1 NCI_CGAP_I
gb_est2:AK008335 + 996.00 1846.67 1.1e-93 ! AK008335 Mus musculus adult mal
gb_hic:AK008335 + 996.00 1846.58 1.1e-93 ! BG244455 602356590P1 NCI_CGAP_M
gb_est2:BG244455 + 991.00 1837.84 3.3e-93 ! BI155774 602904434P1 NIH_CGAP_M
gb_est2:BI155774 + 993.00 1823.56 1.2e-92 ! BI154844 602902802P1 NIH_CGAP_M
gb_est2:BG871384 + 980.00 1818.13 4.2e-91 ! BG871384 602799566P1 NCI_CGAP_S
gb_est2:BR163190 + 969.00 1795.93 7.2e-92 ! BR163190 602717917P1 NCI_CGAP_I
gb_est2:BG6975104 + 967.00 1793.13 1.0e-90 ! BG6975104 602843360P1 NIH_CGAP_M
gb_est2:BG9291306 + 967.00 1788.17 1.9e-90 ! BG9291306 601669059P1 NCI_CGAP_I
gb_est2:BR2913706 + 965.00 1786.66 2.4e-90 ! BR2913706 602961906P1 NCI_CGAP_I
gb_est2:BF30302505 + 934.00 1769.29 2.2e-89 ! BF30302505 602031534P1 NCI_CGAP_S
gb_est2:BR193698 + 925.00 1715.25 2.2e-86 ! BR193698 vj25cl.xl Barstead bc
gb_est1:AA238752 + 917.50 1701.47 1.3e-85 ! AA238752 my35e09.r1 Barstead mc
gb_est1:AA238752 + 897.00 1663.39 1.7e-83 ! AA238752 my35f04.r1 Barstead mc
gb_est1:AA238564 + 847.00 1570.36 1.6e-78 ! AA238564 vj32f09.xl Soares mamu
gb_est1:A57017116 + 839.00 1556.31 1.6e-77 ! A57017116 vj17h04.r1 Barstead mc
gb_est1:AA234993 + 796.00 1474.30 5.9e-73 ! AA234993 SWMFCAL124SK Brugia ma
gb_est1:AA228200 + 795.00 1474.39 5.8e-73 ! AA228200 vj23f09.r1 Barstead bc
gb_est1:BA6819520 + 763.00 1414.11 1.3e-69 ! BA6819520 vs03d09.r1 Barstead bc
gb_est1:AA734805 + 757.50 1404.10 4.8e-69 ! AA734805 vp38b01.r1 Barstead mc
gb_est1:AA597240 + 743.00 1376.77 1.6e-67 ! AA597240 vo28c11.r1 Barstead mc
gb_est1:AA597240 + 718.00 1328.13 8.2e-65 ! AA597240 602830743P1 NCI_CGAP_C
gb_est2:BG2658181 + 697.00 1294.58 6.0e-63 ! BG2658181 vj23f09.r1 Soares mamu
gb_est1:AA499942 + 687.00 1271.81 1.1e-61 ! AA499942 ui-R-EK0-cgh.f-05-0-U1.s
gb_est1:AA543570 + 653.00 1209.55 3.3e-58 ! AA543570 uz23h10.y1 NCI_CGAP_Ma
gb_est2:BF152060 + 630.00 1169.54 5.6e-56 ! BF152060 vs09g08.r1 Barstead mc
gb_est1:AA689633 + 621.00 1148.96 7.8e-55 ! AA689633 226604 MARC 2BOV Bos t
gb_est2:BF076851 + 604.00 1119.27 3.5e-53 ! BF076851 u222h10.y1 NCI_CGAP_Ma
gb_est2:BF153994 + 572.00 1057.16 1.0e-49 ! BF153994 he21d02.xl NCI_CGAP_CM
gb_est1:AA467554 + 504.00 931.29 1.0e-42 ! AA467554 xab3g01.xl NCI_CGAP_CM
gb_est1:AA704595 + 503.50 932.38 9.0e-43 ! AA704595 vj17h10.r1 NCI_CGAP_CM
gb_est1:AA734993 + 500.00 924.82 2.4e-42 ! AA734993 UI-R-EK0-cgh.f-05-0-U1.s
gb_est2:BT291182 - 500.00 924.10 2.6e-42 ! BT291182 UI-R-EK0-cu-f-11-0-U1.s
gb_est1:AA6895213 + 490.00 905.48 2.8e-41 ! AA6895213 mh22a05.r1 Soares mous
gb_est2:BF548605 + 428.00 792.57 5.8e-35 ! BF548605 vo50e07.r1 Soares mous
gb_est1:AA015317 + 386.00 715.78 1.0e-30 ! AA015317 GH07464.Sprime GH Dros
gb_est1:AA616255 + 381.00 699.47 8.5e-30 ! AA616255 SH014764.Sprime GH Dros
gb_est1:AA108979 + 378.00 694.26 1.7e-29 ! AA108979 GH07464.Sprime GH Dros
gb_est1:AA941788 + 378.00 694.26 1.7e-29 ! AA941788 GH07464.Sprime GH Dros

```


Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 682)
 Carninci, P. and Hayashizaki, Y.
 High-efficiency full-length cDNA cloning
 Methods in enzymology. 303, 19-44 (1999)
 99279253
 MEDLINE
 10349636
 PUBMED

2 (bases 1 to 682)
 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
 Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
 Normalization and subtraction of cap-trapper-selected cDNAs to
 prepare full-length cDNA libraries for rapid discovery of new genes
 Genome research. 10 (10), 1617-1630 (2000)
 20499374
 MEDLINE
 11042159
 PUBMED

3 (bases 1 to 682)
 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
 Konno, H., Akiyama, J., Nishi, K., Kitsunai, T., Tashiro, H., Itoh, M.,
 Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,
 Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
 Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M.,
 Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J.,
 Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
 RIKEN integrated sequence analysis (RISA) system--384-format
 sequencing pipeline with 384 multicapillary sequencer
 Genome research. 10 (11), 1757-1771 (2000)
 20530913
 MEDLINE
 11076861
 PUBMED

4 (bases 1 to 682)
 The RIKEN Genome Exploration Research Group Phase II Team and the
 FANTOM Consortium.
 Functional annotation of a full-length mouse cDNA collection
 Nature 409, 685-690 (2001)
 5 (bases 1 to 682)
 Adachi, J., Aizawa, K., Akahira, S., Akimura, T., Aono, H., Arai, A.,
 Arakawa, T., Carninci, P., Fukuda, S., Fukunishi, Y., Furuno, M.,
 Hanagaki, T., Hara, A., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hori, F.,
 Imotani, K., Ishii, Y., Itoh, M., Izawa, M., Kato, H., Kawai, J.,
 Kojima, Y., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T.,
 Miyazaki, A., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Okazaki, Y.,
 Okido, T., Owa, C., Saito, H., Saito, R., Sakai, C., Sakai, K., Sano, H.,
 Sasaki, D., Shibata, K., Shibata, Y., Shinagawa, A., Shiraki, T.,
 Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F.,
 Tanaka, T., Tejima, Y., Toyota, T., Yamamura, T., Yasunishi, A.,
 Yoshida, K., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.
 Direct Submission
 Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of
 Physical and Chemical Research (RIKEN), Laboratory for Genome
 Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
 RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
 Kanagawa 230-0045, Japan (E-mail: genome-res@gs.riken.go.jp,
 URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222,
 Fax: 81-45-503-9216)
 Please visit our web site (http://genome.gsc.riken.go.jp/) for
 further details.

FEATURES
 source
 1..682
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /db_xref="MGI:1345092"
 /db_xref="MGI:1910652"
 /clone="2010107E11"
 /sex="male"
 /tissue_type="small intestine"
 /clone_lib="RIKEN full-length enriched mouse cDNA library"
 /dev_stage="adult"
 38..586
 /note="putative"
 /codon_start=1
 /protein_id="BAB25611.1"
 /db_xref="GI:12842460"
 /translation="MLFACALLALLGLATSCSFYVPRSEWRALPSECSRLGHPVRYV

CDS
 13-FEB-2001
 EST
 688 bp
 mRNA
 NC1_CGAP_Mam1 Mus musculus cDNA clone IMAGE:4485071 5',
 BG244455
 mRNA sequence.
 BG244455
 BG244455.1 GI:12754270

VISHTAGSCFNPDSCFOOARNVOHYHKNELGWCDVAYNELIGEDGHVYEGRGWNIKG
 DHTGPIINPMISIGITPMGNEMDRVPKRALRAALNLLCEGVSRGFLRSNVEVKGHRDV
 QSTLSPGDQLYQVIOQSWEHYRE"
 666..671
 /note="putative"
 682
 /note="putative"
 147 a 217 c 175 g 143 t
 BASE COUNT
 ORIGIN

alignment_scores:
 Quality: 996.00 Length: 182
 Ratio: 5.473 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
 US-09-462-625-2 x AK008335 ..
 Align seg 1/1 to: AK008335 from: 1 to: 682

1 MetLeuPheAlaCysAlaLeuLeuAlaLeuLeuGlyLeuAlaThrSerCy 17
 |||||
 38 ATGTTGTTGGCCTGCTCTCCCTGGCCCTCTGGTGGTGGCAACCTCCG 87
 |||||
 17 sSerPheIleValProArgSerGluTrpArgAlaLeuProSerGluCys 34
 |||||
 88 CAGTTTCATCGTGGCCCGCAGTGGAGTGGAGGCCCTGCCATCGAGTGCT 137
 |||||
 34 erSerArgLeuGlyHisProValArgTrpValValIleSerHisThrAla 50
 |||||
 138 CTAGCGCCTGGGGCCACCCAGTTTCGCTACGTGGTGTATCTCACACAGCC 187
 |||||
 51 GlySerPheCysAsnSerProAspSerCysGluGlnGlnAlaArgAsnVa 67
 |||||
 188 GGCAAGTTTCGCAACACCCCGGACTCTCTGGAACAGCAGGCCCGCATGT 237
 |||||
 67 lGlnHisTyrHisLysAsnGluLeuGlyTrpCysAspValAlaTyrAsn 84
 |||||
 238 GCAGCATTTACCACAAGAATCAGCTGGCTGGTGGTGGTGGTGGTGGT 287
 |||||
 84 heLeuIleGlyGluAspGlyHisValTyrGluGlyArgGlyTrpAsnIle 100
 |||||
 288 TCCTTATTTGAGAGGAGCGGTCTCTATGAAGCCGAGGCTGGAACATC 337
 |||||
 101 LysGlyAspHisThrGlyProIleTrpAsnProMetSerIleGlyIleTh 117
 |||||
 338 AAGGTGACCCACACAGGCCCATCTGGAATCCCATGCTATTGGCATCAC 387
 |||||
 117 rPheMetGlyAsnPheMetAspArgValProAlaLysArgAlaLeuArg 134
 |||||
 388 CTTTCATGGGAACCTTCATGGACCGGGTACCCGCAAGCGGCCCTCCGTG 437
 |||||
 134 laAlaLeuAsnLeuLeuGluCysGlyValSerArgGlyPheLeuArgSer 150
 |||||
 438 CTGCCCTAAATCTTCTGGAATGTGGGTGTCTCGGGCTTCTTGAGATCC 487
 |||||
 151 AsnTyrGluValLysGlyHisArgAspValGlnSerThrLeuSerProGl 167
 |||||
 488 AACTATCAAGTCAAGGACACCGGGATGTCAAGAGCACTCTCTCTCCAGG 537
 |||||
 167 YAspGlnLeuTyrGlnValIleGlnSerTrpGluHisTyrArgGlu 182
 |||||
 538 TGACCAACTATCAGGTCTATCCAAAGCTGGGAACACTACCGAGAG 583
 |||||

seq_name: gb_est2:BG244455

seq_documentation_block:
 LOCUS BG244455 688 bp mRNA
 DEFINITION 602356590F1 NC1_CGAP_Mam1 Mus musculus cDNA clone IMAGE:4485071 5',
 BG244455
 mRNA sequence.
 ACCESSION BG244455
 VERSION BG244455.1 GI:12754270

```

|||||
319 AAGGTGACACACAGGGCCCATCTGGAATCCCATGCTATTGGCATCAC 368
|||||
117 rPheMetGlyAsnPhetMetAspArgValProAlaLysArgAlaLeuArgA 134
|||||
369 CTTCATGGGGAACCTTCATGGACCGGTACCCGCAAGCGGCCCTCCGTG 418
|||||
134 laAlaLeuAsnLeuLeuGluCysGlyValSerArgGlyPheLeuArgSer 150
|||||
419 CTGCCCTAAATCTTCTGGAATGGGGTGTCTCGGGGCTCTCTGAGATCC 468
|||||
151 AsnTyrGluValLysGlyHisArgAspValGlnSerThrLeuSerProG1 167
|||||
469 AACTATGAAGTCAAAAGGACACCGGATGTGCAAGACACTCTCTCTCCAGG 518
|||||
167 yAspGlnLeuTyrGlnValIleGlnSerTyrGlnHisTyrArgGlu 182
|||||
519 TGACCAACTCTATCAGGTCTATCAAGCTGGGACACTACCGAGAG 564
|||||

```

seq_name: gb_est2:BI155774

seq_documentation_block: 650 bp mRNA EST 05-JUL-2001
LOCUS BI155774 NIH_CGAP_Mam3 Mus musculus cDNA clone IMAGE:5033867 5',
DEFINITION mRNA sequence.

ACCESSION BI155774
VERSION BI155774.1 GI:14615775

KEYWORDS EST.
SOURCE house mouse.

ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 650)

AUTHORS NIH-MGC http://mgc.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
Tissue Procurement: Lothar Hennighausen Ph.D., Chu-Xia Deng Ph.D.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM1094 row: f column: 12
High quality sequence stop: 637.

FEATURES
Location/Qualifiers

1..650
source
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:5033867"
/clone_lib="NIH_CGAP_Mam3"
/tissue_type="tumor, gross tissue"
/lab_host="DH10B"
/note="Organ: mammary; Vector: pCMV-SPORT6; Site:1: NotI;
Site:2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert 2 kb. Library constructed by Life
Technologies, catalog #12017-018. Investigators providing
samples: Lothar Hennighausen/Chu-Xia Deng, NIH Reference
for transgenic model: Xu et al., Nature Genetics 22, 37-43
(1999). Note: this is a NCI_CGAP Library."

BASE COUNT 137 a 199 c 179 g 135 t
ORIGIN

alignment_scores:
Quality: 991.00 Length: 182
Ratio: 5.445 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 99.451

alignment_block:
US-09-462-625-2 x BI155774

KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE NIH-MGC http://mgc.nci.nih.gov/
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
Tissue Procurement: Lothar Hennighausen Ph.D.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM10326 row: c column: 24
High quality sequence stop: 587.

FEATURES
Location/Qualifiers

1..688
source
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:4485071"
/clone_lib="NCI_CGAP_Mam1"
/tissue_type="tumor, biopsy sample"
/dev_stage="3 months, virgin"
/lab_host="DH10B"
/note="Organ: mammary; Vector: pCMV-SPORT6; Site:1: SalI;
Site:2: NotI; Cloned unidirectionally. Primer: Oligo dT.
Library constructed by Life Technologies. Investigator
providing samples: Gilbert Smith, NIH"
BASE COUNT 157 a 209 c 180 g 142 t
ORIGIN

alignment_scores:
Quality: 996.00 Length: 182
Ratio: 5.473 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-462-625-2 x BG244455

Align seg 1/1 to: BG244455 from: 1 to: 688

```

1 MetLeuPheAlaCysAlaLeuLeuAlaLeuGlyLeuAlaThrSerCy 17
|||||
19 ATGTTGTTGCTGCTGCTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCT 68
|||||
17 sSerPheIleValProArgSerGluTyrArgAlaLeuProSerGluCys 34
|||||
69 CAGTTTCATGTCGCCCGCAGTGAAGGGCCCTGCCATCCGAGTGCT 118
|||||
34 erSerArgLeuGlyHisProValArgTyrValValIleSerHisThrAla 50
|||||
119 CTAGCCGCTGGGGACCCAGTTCGCTACGTGCTGATCTCACACAGGCC 168
|||||
51 GlySerPheCysAsnSerProAspSerCysGluGlnAlaArgAsnVa 67
|||||
169 GGCAGCTTCTGCAACAGCCGCGACTCTCTGTAACAGCAGCGCCGCAATGT 218
|||||
67 IGLnHisTyrHisLysAsnGluLeuGlyTyrCysAspValAlaTyrAsnP 84
|||||
219 GCAGCATTAACCAAGATGACGTGGCTGGCGATGTAGCTACAAC 268
|||||
84 heLeuIleGlyGluAspGlyHisValTyrGluGlyArgGlyTyrAsnIle 100
|||||
269 TCCTTATTGGAGGACGCTCATGCTATGAAGCGCGAGCGGCGAATC 318
|||||
101 LysGlyAspHisThrGlyProIleTyrPasnProMetSerIleGlyIleTh 117
|||||

```

Align seg 1/1 to: B1155774 from: 1 to: 650

```

1 MetLeuPheAlaCysAlaLeuLeuAlaLeuGlyLeuAlaThrSerCy 17
|||||
71 ATGTTGTTTGGCTGCTGCTCTCTCCCTCTCTGGTCTGGCAACCTCTG 120
|||||
17 sSerPheIleValProArgSerGluTrpArgAlaLeuProSerGluCys 34
|||||
121 CAGTTTCATCGTGGCCCCGAGTGAAGGCGCTGCCATCCGAGTCT 170
|||||
34 eSerArgLeuGlyHisProValArgTyrValValIleSerHisThrAla 50
|||||
171 CTAGCGCGCTGGGGCACCAGTTCGTACGTGCTGATCTCACACAGCC 220
|||||
51 GlySerPheCysAsnSerProAspSerCysGluGlnGlnAlaArgAsnVa 67
|||||
221 GGCAGCTTCTGCACAGCCCGGACCTCTGTGAACAGCAGGCGCCGCAATGT 270
|||||
67 lGlnHisTyrHisLysAsnGluLeuGlyTrpCysAspValAlaTyrAsnP 84
|||||
271 GCAGCATTAACACACAGATGAGTGGCTGGTGGCATGTAGCTACAACT 320
|||||
84 heLeuIleGlyClnAspGlyHisValTyrGluGlyArgGlyTrpAsnIle 100
|||||
321 TCCTTATTGGAGAGCGTCTATGTAAGCGCGAGGCTGGAACATC 370
|||||
101 LysGlyAspHisThrGlyProIleTrpAsnProMetSerIleGlyIleTh 117
|||||
371 AAGGTGTACACACAGGCGCCATCTGGAATCCATCTATTGGCATCAC 420
|||||
117 rPheMetGlyAsnPheMetAspArgValProAlaLysArgAlaLeuArgA 134
|||||
421 CTTCTATGGGAACTTCATGGACCGGTACCCGCCAAGCGGGCCCTCCGTG 470
|||||
134 laAlaLeuAsnLeuLeuGlyCysGlyValSerArgGlyPheLeuArgSer 150
|||||
471 CTGCGCTACTCTTCTGGAATGGGGTGTCTCGGGGCTTCTTGAGATCC 520
|||||
151 AsnTyrGluValLysGlyHisArgAspValGlnSerThrLeuSerProGl 167
|||||
521 AACTATGAAGTCAAAAGGACACCGCGGATGTGCAAGCACTCTCTCCAGG 570
|||||
167 yAspGlnLeuTyrGlnValIleGlnSerTrpGluHisTyrArgGlu 182
|||||
571 TGACCAACTCTATCAGGTCTATCAAGTGGGAACACATCCAGAGAG 616

```

seq_name: gb_est2:B1154844

seq_documentation_block: 610 bp mRNA EST 05-JUL-2001
LOCUS B1154844 602902826F1 NIH_CGAP_Mam3 Mus musculus cDNA clone IMAGE:5032278 5',
DEFINITION

mRNA sequence.

ACCESSION B1154844

VERSION B1154844.1 GI:14614845

KEYWORDS EST.

SOURCE house mouse.

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

REFERENCE

AUTHORS NIH-MGC http://mgi.nci.nih.gov/.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgaps-r@mail.nih.gov

Tissue Procurement: Lothar Hennighausen Ph.D., Chu-Xia Deng Ph.D.

cDNA Library Preparation: Life Technologies, Inc.

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LLAM11090 row: d column: 07
High quality sequence stop: 610.

FEATURES

source

1..610

/organism="Mus musculus"

/db_xref="taxon:10090"

/clone="IMAGE:5032278"

/clone_lib="NIH_CGAP_Mam3"

/tissue_type="tumor, gross tissue"

/lab_host="DH10B"

/note="Organ: mammary; Vector: pCMV-SPORT6; Site: 1: NotI;
Site: 2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert 2 kb. Library constructed by Life
Technologies, catalog #12017-018. Investigators providing
samples: Lothar Hennighausen/Chu-Xia Deng, NIH Reference
for transgenic model: Xu et al., Nature Genetics 22, 37-43
(1999). Note: this is a NCI-CGAP Library."

BASE COUNT 128 a 179 c 173 g 130 t
ORIGIN

alignment_scores:

Quality: 983.00 Length: 183

Ratio: 5.401 Gaps: 1

Percent Similarity: 99.454 Percent Identity: 99.454

alignment_block:

US-09-462-625-2 x B1154844

Align seg 1/1 to: B1154844 from: 1 to: 610

```

1 MetLeuPheAlaCysAlaLeuLeuAlaLeuGlyLeuAlaThrSerC 17
|||||
31 ATGTTGTTTGGCTGCTGCTCTCTCCCTCTCTGGTCTGGCAACCTCT 80
|||||
17 ySerPheIleValProArgSerGluTrpArgAlaLeuProSerGluCys 33
|||||
81 GCAGTTTCATCGTGGCGCGAGTGAAGGCGGCGCTGCCATCCGAGTGC 130
|||||
34 SerSerArgLeuGlyHisProValArgTyrValValIleSerHisThrAl 50
|||||
131 TCTAGCGCCCTGGGGCCACCCAGTTCGCTACGTGGTGTGATCTCACACAGC 180
|||||
50 aGlySerPheCysAsnSerProAspSerCysGluGlnGlnAlaArgAsnV 67
|||||
181 CGGACAGTCTTGCAACAGCGCGGACTCTCTGTGAACAGCAGCGCCGCAATG 230
|||||
67 aGlnHisTyrHisLysAsnGluLeuGlyTrpCysAspValAlaTyrAsn 83
|||||
231 TGCAGCATTAACACACAGATGAGTGGGCTGGTGGCATGTAGCTTACACAC 280
|||||
84 PheLeuIleGlyGluAspGlyHisValTyrGluGlyArgGlyTrpAsnI 100
|||||
281 TTCTTATTGGAGAGCGGTCATGCTATGAAGCGCGGAGGCTGGAACAT 330
|||||
100 eLysGlyAspHisThrGlyProIleTrpAsnProMetSerIleGlyIleT 117
|||||
331 CAAGGTCACACACAGCGGCCCATCTGGATCCCATGCTATTGGCATCA 380
|||||
117 hrPheMetGlyAsnPheMetAspArgValProAlaLysArgAlaLeuArg 133
|||||
381 CTTTCATGGGAACATTCATGACCGGGTACCGCAAGCGGCGCTCCGT 430
|||||
134 AlaAlaLeuAsnLeuLeuGluCysGlyValSerArgGlyPheLeuArgSe 150
|||||
431 GCTGCCCTAAATCTTCTGGAATGTGGGTGCTCGGGGCTTCTTGAGATC 480
|||||
150 rAsnTyrGluValLysGlyHisArgAspValGlnSerThrLeuSerProG 167
|||||
481 CAACATATGAAGTCAAGGACACCGGAGATGCAAGAGCACTCTCTCTCCAG 530
|||||
167 lyAspGlnLeuTyrGlnValIleGlnSerTrpGluHisTyrArgGlu 182
|||||

```

```

531 GTGACCAACTCTATCAGTCTATCAAGAGCTGGGAACACTACCGAGAG 577
seq_name: gb_est2:BG871384

seq_documentation_block:
LOCUS      BG871384          600 bp      mRNA          EST          29-MAY-2001
DEFINITION 602790566F1 NCI_CGAP_SG2 Mus musculus cDNA clone IMAGE:4921697 5',
mRNA sequence.
ACCESSION  BG871384
VERSION     BG871384
KEYWORDS    BG871384.1 GI:14221924
SOURCE      house mouse.
ORGANISM    Mus musculus
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE   1 (bases 1 to 600)
AUTHORS     NIH-MGC http://mgc.nci.nih.gov/
TITLE       National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL     Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Jeffrey E. Green, M.D.
            cDNA Library Preparation: Life Technologies, Inc.
            CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10840 row: d column: 18
            High quality sequence stop: 585.
            Location/Qualifiers
              1..600
                /organism="Mus musculus"
                /strain="FVB/N"
                /db_xref="taxon:10090"
                /clone="IMAGE:4921697"
                /clone_lib="NCI_CGAP_SG2"
                /lab_host="DH10B (TI phase-resistant)"
                /note="Organ: salivary gland; Vector: pCMV-SPORT6; Site_1:
                NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo
                dt. Average insert size 1.3 kb. Constructed by Life
                Technologies. Note: this is a NCI_CGAP Library."

BASE COUNT      126 a 178 c 167 g 129 t
ORIGIN

alignment_scores:
  Quality: 98.00      Length: 183
  Ratio: 5.385       Gaps: 1
  Percent Similarity: 99.454      Percent Identity: 98.907

alignment_block:
US-09-462-625-2 x BG871384 ..
Align seg 1/1 to: BG871384 from: 1 to: 600

1 MetLeuPheAlaCysAlaLeuLeuAlaLeuLeuGlyLeuAlaThrSerCy 17
|||||
18 ATCTGTTTCCCTGTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 67
|||||
17 sSerPheIleValProHrgSerGluTrpArgAlaLeuProSerGluCysS 34
|||||
68 CAGTTTCATCGTCCCGCCAGTGAAGTGGAGGCGCCCTGCCATCCGAGTGT 117
|||||
34 eSerArgLeuGlyHisProValArgTyrValValIleSerHisThrAla 50
|||||
118 CTAGCCCGCTGGGACCCAGTTCCTCTCTCTCTCTCTCTCTCTCTCTCTCT 167
|||||
51 GlySerPheCysAsnSerProAspSerCysGluGlnGlnAlaArgAsnVa 67
|||||
168 GGCAGCTTCTGCAACAGCCCGGACTCTCTGTGAACAGCAGCGCCGCAATGT 217
|||||
67 lGlnHisTyrHisLysAsnGluLeuGlyTyrCysAspValAlaTyrAsnP 84

```

```

|||||
218 GCAGCATTACCAAGAAGTGAAGCTGGCTGGTGGATGTAGCTACACT 267
|||||
84 heLeuIleGlyGluAspGlyHisValTyrGluGlyArgGlyTyrAsnIle 100
|||||
268 TCCTTATTTGAGAGGACGGTCAATGCTATGAAGCGCGAGGCTGGACATC 317
|||||
101 LysGlyAspHisThrGlyProIleTyrPAsnProMetSerIleGlyIleTh 117
|||||
318 AAGGTTGACCACACAGGCGCCATCTGGAATCCCATGTCTATTTGGCATCAC 367
|||||
117 rPheMetGlyAsnPheMetAspArgValPrcAlaLysArgAlaLeuArgA 134
|||||
368 CTTTCATGGGGAACCTCATGACCGGGTACCGCAAGGCGCCCTCCGTG 417
|||||
134 laAlaLeuAsnLeuLeuGlyCysGlyValSerArgGlyPheLeuArgSer 150
|||||
418 CTGCCCTAAATCTTCTGAAATGTGGGTGTCTCGGGGCTTCTTGAGATCC 467
|||||
151 AsnTyrGluValLysGlyHisArgAspValGlnSerThrLeuSerProGl 167
|||||
468 AACTATGAAGTCAAGGACACCGGGATGTGCAAGACACTCTCTCTCCAGG 517
|||||
167 YAspGlnLeuTyrGlnValIleGlnSer.TrpGluHisTyrArgGlu 182
|||||
518 TGACCAACTCTATCAGTCTATCAACAAACGCTGGGAACACTACCGAGAG 564
|||||
seq_name: gb_est2:BF163190

seq_documentation_block:
LOCUS      BF163190          706 bp      mRNA          EST          30-OCT-2000
DEFINITION 601771917F1 NCI_CGAP_Lu29 Mus musculus cDNA clone IMAGE:3990820 5',
mRNA sequence.
ACCESSION  BF163190
VERSION     BF163190.1 GI:11043436
KEYWORDS    BF163190
SOURCE      house mouse.
ORGANISM    Mus musculus
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE   1 (bases 1 to 706)
AUTHORS     NIH-MGC http://mgc.nci.nih.gov/
TITLE       National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL     Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Gilbert Smith, Ph.D.
            CDNA Library Preparation: Life Technologies, Inc.
            CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM9203 row: b column: 05
            High quality sequence stop: 622.
            Location/Qualifiers
              1..706
                /organism="Mus musculus"
                /strain="C57BL/6J (fetal)"
                /db_xref="taxon:10090"
                /clone="IMAGE:3990820"
                /clone_lib="NCI_CGAP_Lu29"
                /tissue_type="spontaneous tumor, metastatic to mammary.
                Stem cell origin."
                /lab_host="DH10B"
                /note="Organ: lung; Vector: pCMV-SPORT6; Site_1: SalI;
                Site_2: NotI; Cloned unidirectionally. Primer: Oligo dt.
                Library constructed by Life Technologies. Investigator
                providing samples: Gilbert Smith, NIH"

BASE COUNT      160 a 213 c 187 g 146 t
ORIGIN

```



```

134 laalaleuasnleuLeuGluCysGlyValSerArgGlyPheLeuArgSer 150
|||||
466 CTGCCCTAAATCTCTGGGAATGTTGGGTGCTCGGGCTTCCTGAGATCC 515
|||||
151 AsnTyrGluValLysGlyHisArgAspValGlnSerThrLeuSerProG1 167
|||||
516 AACTATGACGTCAAGGACACCGGGATGTGCAAGCACTCTCTCTCAGG 565
|||||
167 yAspGlnLeuTyrGlnValLleGln.SerTrpGluHisTyrArgGlu 182
|||||
566 TGACCAACTCTATCAGGTCTATCCCAACCGCTGGGAACACTACCGGAG 612
|||||

seq_name: gb_est2:BE913706

seq_documentation_block:
LOCUS BE913706 1045 bp mRNA EST 29-SEP-2000
DEFINITION 601669059F1 NCI_CGAP_Mam1 Mus musculus cDNA clone IMAGE:3969014 5',
mRNA sequence.
ACCESSION BE913706
VERSION BE913706.1 GI:10411594
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE 1 (bases 1 to 1045)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1AM9146 row: e column: 15
High quality sequence stop: 618.

FEATURES
source
location/Qualifiers
1..1045 Mus musculus
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:3969014"
/clone_lib="NCI_CGAP_Mam1"
/tissue_type="tumor, biopsy sample"
/dev_stage="3 months, virgin"
/lab_host="DH10B"
/note="Organ: mammary; Vector: pCMV-SPORT6; Site:1: SalI;
Site:2: NotI; Cloned unidirectionally. Primer: Oligo dt.
Library constructed by Life Technologies. Investigator
providing samples: Gilbert Smith, NIH"
BASE COUNT 315 a 305 c 261 g 164 t
ORIGIN

alignment_scores:
Quality: 967.00 Length: 184
Ratio: 5.313 Gaps: 2
Percent Similarity: 98.913 Percent identity: 98.370

alignment_block:
US-09-462-625-2 x BE913706 ..
Align seg 1/1 to: BE913706 from: 1 to: 1045
1 MetLeuPheAlaCysAlaLeuLeuAlaLeuGlyLeuAlaThrSerCys 17
|||||
51 ATGTTGTTTGCCTGCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 100
|||||
17 sserPheIleValProArgSerGluTrpArgAlaLeuProSerGluCys 34
|||||

```

```

|||||
101 CAGTTTCATCGTCCCGCCAGTGAAGTGGGCGCCCTGCCATCCGAGTGT 150
|||||
34 erSerArgLeuGlyHisProValArgTyrValValIleSerHisThrAla 50
|||||
151 CTAGCGGCTGGGGCACCAGTTCGCTACGTGCTGATCTCACACACAGCC 200
|||||
51 GlySerPheCysAsnSerProAspSerCysGluGlnGlnAlaArgAsnVa 67
|||||
201 GGCAGCTTCTGCAACAGCCCGGACTCTGTGAACAGCAGGCGCGCAATGT 250
|||||
67 GlnHisTyrHisLysAsnGluLeuGlyTrpCysaspValAlaTyrAsnP 84
|||||
251 GCAGCATTAACCAAGATGAGCTGGGCTGGTGGATCCATGCTATGGCATC 300
|||||
84 heLeuIleGlyGluAspGlyHisValTyrGluGlyArgGlyTrpAsnIle 100
|||||
301 TCCTTTATGGAGAGGAGCGGTCACTGTATGAAGCGCGGAGGTGAACATC 350
|||||
101 LysGlyAspHisThrGlyProIleTrpAsnPrometSerIleGlyIleTh 117
|||||
351 AAGGGTGACCAACAGGCGCCATCTGGAATCCATGCTATGGCATCAC 400
|||||
117 rPheMetGlyAsnPheMetaspArgValProAlaLysArgAlaLeuArgA 134
|||||
401 CTTTCATGGGAACTCTATGACCGGGTACCCTGCAAGCGGCGCTCCGTG 450
|||||
134 laalaleuasnleuLeuGluCysGlyValSerArgGlyPheLeuArgSer 150
|||||
451 CTGCCCTAAATCTCTCTGGAATGTTGGGTGCTCGGGCTTCCTGAGATCC 500
|||||
151 AsnTyr.GluValLysGlyHisArgAsp.ValGlnSerThrLeuSerPro 166
|||||
501 AACATATGCAAGTCAAAAGACACCGGGATGTGCAAGCACTCTCTCTCCA 550
|||||
167 GlyAspGlnLeuTyrGlnValLleGlnSerTrpGluHisTyrArgGlu 182
|||||
551 GGTGACCAACTCTATCAGGTCTATCCCAACCGTGGGAACACTACCGGAG 598
|||||

seq_name: gb_est2:BI409815

seq_documentation_block:
LOCUS BI409815 842 bp mRNA EST 14-AUG-2001
DEFINITION 602961906F1 NCI_CGAP_Lu33 Mus musculus cDNA clone IMAGE:5117688 5',
mRNA sequence.
ACCESSION BI409815
VERSION BI409815.1 GI:15170738
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE 1 (bases 1 to 842)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1AM1286 row: k column: 01
High quality sequence stop: 568.

FEATURES
source
location/Qualifiers
1..842 Mus musculus
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"

```

```

/clone="IMAGE:5117688"
/clone_lib="NCI_CGAP_Lu33"
/tissue_type="pooled lung tumors"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: lung; Vector: pT7T3D-Pac (Pharmacia) with a
modified polylinker; Site.1: NotI; Site.2: EcoRI; 1st
strand cDNA was prepared from mRNA obtained from pooled
lung tumors with a Not I - oligo(dT) primer [5'
TGTACCAATCTGAAGTGGAGCGCGCCTCTGTTTTTTTTTTT 3'].
Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not
I and Eco RI sites of the modified pT7T3 vector. Library
went through one round of normalization, and was
constructed by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT 201 a 259 c 226 g 154 t 2 others
ORIGIN

alignment_scores:
  Quality: 965.00 Length: 178
  Ratio: 5.483 Gaps: 0
Percent Similarity: 98.876 Percent Identity: 98.315

alignment_block:
US-09-462-625-2 x B1409815 ..
Align seg 1/1 to: B1409815 from: 1 to: 842
5 CysAlaLeuAlaLeuLeuGlyLeuAlaThrSerCysSerPheIleVa 21
|||||
3 TGTGCTCTCTCGCCCTTATGGTCTGGCAACCTCTCGAGTTTCATCGT 52
|||||
21 lProArgSerGluTrpArgAlaLeuProSerGluCysSerArgLeuG 38
|||||
53 GCGCCCGCAGTAGTGGAGGCGCTGCCATCCGAGTGTCTAGCGCGCTGG 102
|||||
38 lyHisProValArgTyrValIleSerHisThrAlaGlySerPheCys 54
|||||
103 GGCACCCAGCTTCGCTACGTGGTGTATCTCACACAGCGCGCAGCTTCTGC 152
|||||
55 AsnSerProAspSerCysGluGlnGlnAlaArgAsnValGlnHisTyrHI 71
|||||
153 AACAGCCCGGACTCTCTGTGAACAGCAGGCGCCGCAATGTGCAGCATTAACA 202
|||||
71 sLysAsnGluLeuGlyTrpCysAspValAlaTyrAsnPheLeuIleGlyG 88
|||||
203 CAGAATAGCTGGGTGGTGGCATGTAGCTACAACTTCTTANTGGAG 252
|||||
88 luAspGlyHisValTyrGluGlyArgGlyTrpAsnIleLysGlyAspHis 104
|||||
253 AGGAGGTCATGTCTATGAAGCGCGAGGCTGGAACATCAAGGCTGACCAC 302
|||||
105 ThrGlyProIleTrpAsnProMetSerIleGlyIleThrPheMetGlyAs 121
|||||
303 ACAGGGCCCATCTGGAAATCCATGTCTANTGGCATCACCTTCATGGGAA 352
|||||
121 nPheMetAspArgValProAlaLysArgAlaLeuArgAlaAlaLeuAsnL 138
|||||
353 CTTTCATGGACCGGTATCCCAAGGGGCGCTCCGTCGTCCTCCCTTAATC 402
|||||
138 euLeuGluCysGlyValSerArgGlyPheLeuArgSerAsnTyrGluVal 154
|||||
403 TTCTGGAATCTGGGTGTCTCGGGGCTCTCTGAGATCCAACCTATGAATC 452
|||||
155 LysGlyHisArgAspValGlnSerThrLeuSerProGlyAspGlnLeuTy 171
|||||
453 AAAGGACACCGGATGTGCAAGACACTCTCTCCAGGTGACCAACTCTA 502
|||||
171 rGlnValIleGlnSerTrpGluHisTyrArgGlu 182
|||||
503 TCAGGTATCAAGCTGGGAACACTACCCAGAG 536
|||||
seq_name: gb_est2:BF302505

```

```

seq_documentation_block:
LOCUS BF302505 619 bp mRNA EST 21-NOV-2000
DEFINITION 602031534F1 NCI_CGAP_SG2 Mus musculus cDNA clone IMAGE:4166360 5',
mRNA sequence.
ACCESSION BF302505
VERSION BF302505.1 GI:11249045
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus.
REFERENCE 1 (bases 1 to 619)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
AUTHORS NTH-MGC http://mgc.nci.nih.gov/.
TITLE NTH-MGC
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM9455 row: d column: 09
High quality sequence stop: 607.
FEATURES
  source
    1..619
    /organism="Mus musculus"
    /strain="FVB/N"
    /db_xref="taxon:10090"
    /clone="IMAGE:4166360"
    /clone_lib="NCI_CGAP_SG2"
    /lab_host="DH10B (TI phage-resistant)"
    /note="Organ: salivary gland; Vector: pCMV-SPORT6; Site.1:
    NotI; Site.2: SalI; Cloned unidirectionally. Primer: Oligo
    dt. Average insert size 1.3 kb. Constructed by Life
    Technologies. Note: this is a NCI_CGAP Library."
  BASE COUNT 132 a 185 c 169 g 133 t
  ORIGIN

```

```

alignment_scores:
  Quality: 954.00 Length: 179
  Ratio: 5.330 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 99.441

alignment_block:
US-09-462-625-2 x BF302505 ..
Align seg 1/1 to: BF302505 from: 1 to: 619
4 AlaCysAlaLeuLeuAlaLeuLeuGlyLeuAlaThrSerCysSerPheIle 20
|||||
52 GCCTGTGCTCTCTCTGCTCTGCTGGTCTGGCAACCTCTCTGAGTTTCAT 101
|||||
20 eValProArgSerGluTrpArgAlaLeuProSerGluCysSerArgL 37
|||||
102 CGTGGCCCGCAGTAGTGGAGGCGCTGCCATCCGAGTGTCTAGCGGCC 151
|||||
37 euGlyHisProValArgTyrValIleSerHisThrAlaGlySerPhe 53
|||||
152 TGGGGCACCCAGTTCGCTACGTGGTATCTCACACAGCGCGCAGCTTC 201
|||||
54 CysAsnSerProAspSerCysGluGlnGlnAlaArgAsnValGlnHisTy 70
|||||
202 TGCAACAGCGCGGACTCTGTGAACAGCAGCGCGCCGCAATGTGCAGCATTA 251
|||||
70 rHisLysAsnGluLeuGlyTrpCysAspValAlaTyrAsnPheLeuIleG 87
|||||
252 CCACAAGATGAGCTGGGCTGGTGGATGTAGCTACAACTTCTCTTAT.G 300
|||||

```


87 lyGluaspGlyHisValTyrGluGlyArgGlyTrpAsnIleLysGlyAsp 103
|||||
301 GAGAGGACGGTCACTGTATGAAGCGCGAGGCTGGAAACATCAAGGGTGAC 350
104 HisThrGlyProIleTyrAsnProMetSerIleGlyIleThrPheMetG1 120
|||||
351 CACACAGGGCCCATCTGGATCCATGCTCTATTGGCATCACCTTCATGGG 400
120 yAsnPheMetAspArgValProAlaLysArgAlaLeuArgAlaLeuA 137
|||||
401 GAACCTTCATGACCGGCTACCGCAAGCGGGCCCTCGGTGCTGCCCTAA 450
137 snLeuLeuGluCysGlyValSerArgGlyPheLeuArgSerAsnTyrGlu 153
|||||
451 ATCTTCTGGAATGTGGGTGTCTCGGGCTTCTTGGATCCCAATATGAA 500
154 ValLysGlyHisArgAspValGlnSerThrLeuSerProGlyAspGlnLe 170
|||||
501 GTCAAGGACACCGGAGTGTCAAGACACTCTCTCTCAGGTGACCAACT 550
170 uTyrGlnValIleGlnSerTrpGluHisTyrArgGlu 182
|||||
551 CTATCAGGTATCCAAAGCTGGGAACACTACCGAGAG 587

seq_name: gb_est1:BE199698

seq_documentation_block: 614 bp mRNA EST 26-JUN-2000
LOCUS BE199698
DEFINITION ug52c11.x1 Barstead bowel MPLRB9 Mus musculus cDNA clone
IMAGE:1546004 3' similar to gb:X86374 M.musculus mRNA for TAG7
protein (MOUSE);, mRNA sequence.
ACCESSION BE199698
VERSION BE199698.1 GI:8711867
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 614)
AUTHORS Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,
Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person
B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter
E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,
Waterston,R. and Wilson,R.
The WashU-NCI Mouse EST Project 1999
Unpublished (1999)
Other_ESTs: ug52c11.y1
Contact: Marra M/WashU-NCI Mouse EST Project 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
MGI:951352
High quality sequence stop: 451.
Location/Qualifiers
1..614
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:1546004"
/clone.lib="Barstead bowel MPLRB9"
/tissue_type="bowel"
/dev_stage="8 weeks"
/lab_host="DH10B"
/note="Vector: p7T3D-Pac (Pharmacia) with a modified
polylinker; Site_1: EcoRI; Site_2: NotI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5',
TGTTACGACTGTAAGTGGGCGCGCCCTTTTTTTTTTTTTTTTTTT
3']; double-stranded cDNA was ligated to Eco RI adaptors
[AATTGGATCCCTTG], digested with Not I and cloned into the

Not I and Eco RI sites of the modified p7T3 vector.
Source irradiated bowel harvested 72 hours after
irradiation (1400 Gys). Library constructed by Bob.
Barstead.

BASE COUNT 128 a 159 c 187 g 139 t 1 others
ORIGIN

alignment_scores:
Quality: 925.00 Length: 173
Ratio: 5.441 Gaps: 0
Percent Similarity: 98.266 Percent Identity: 97.688

alignment_block:
US-09-462-625-2 x BE199698/rev ..
Align seg 1/1 to reverse of: BE199698 from: 1 to: 614
10 LeuLeuGlyLeuAlaThrSerCysSerPheIleValProArgSerGluTr 26
|||||
614 CTCITGGTGGCGCAACCTCTGCGCAGTTTCATCGCCCCGCGAGTG 565
26 pArgAlaLeuProSerGluCysSerSerArgLeuGlyHisProValArgT 43
|||||
564 GAGGCGCTTGCCATCCGAGTGTCTAGCCGCTTGGGGCACCGACTTCGCT 515
43 yrValValIleSerHisThrAlaGlySerPheCysAsnSerProAspSer 59
|||||
514 ACGRGTGTATCTCACACACCGCGCAGCTTNTGCACAGCCCGGACTCC 465
60 CysGluGlnGlnAlaArgAsnValGlnHisTyrHisLysAsnGluLeuG1 76
|||||
464 TGTGAACAGCAGCGCCGCAATGTGCAGCATACCACAGAATGACCTGGG 415
76 yTrpCysAspValAlaTyrAsnPheLeuIleGlyGluAspGlyHisValT 93
|||||
414 CTGGTGGCATGTAGCTCAACTTCTTATTGGAGAGACGGTCACTGCT 365
93 yrGluGlyArgGlyTrpAsnIleLysGlyAspHisThrGlyProIleTrp 109
|||||
364 ATGAAGCGCGAGGCTGGAAACATCAAGGGGTGACACACAGGGCCCATCTGG 315
110 AsnProMetSerIleGlyIleThrPheMetGlyAsnPheMetAspArgVa 126
|||||
314 ATCCCATGTCTATTGGCATCATCTTCATGGGGAACCTCATGACCGGCT 265
126 lProAlaLysArgAlaLeuArgAlaAlaLeuAsnLeuLeuGluCysGlyV 143
|||||
264 ACCCGCAAGCGGGCCCTCCGTGCTGCCCTAAATCTTCTGGAATGTGGG 215
143 alSerArgGlyPheLeuArgSerAsnTyrGluValLysGlyHisArgAsp 159
|||||
214 TGTCTCGGGGCTCTTGGATCCACTATGAAGTCAAGGACACCGGGAT 165
160 ValGlnSerThrLeuSerProGlyAspGlnLeuTyrGlnValIleGlnSe 176
|||||
164 GTGCAAGACACTCTCTCTCCAGGTGACCAACTCTATCAGTCAATCCAAAG 115
176 rTrpGluHisTyrArgGlu 182
|||||
114 CTGGGAACACTACCGAGAG 96

seq_name: gb_est1:AA238752

seq_documentation_block: 601 bp mRNA EST 03-MAR-1997
LOCUS AA238752
DEFINITION mv35e09.r1 Barstead mouse pooled organs MPLRB4 Mus musculus CDNA
clone IMAGE:697864 5' similar to gb:X86374 M.musculus mRNA for TAG7
protein (MOUSE);, mRNA sequence.
ACCESSION AA238752
VERSION AA238752.1 GI:1862775
KEYWORDS EST.
SOURCE house mouse.

```

ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 601)
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.
TITLE The WashU-HHMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT Contact: Marra M/Mouse EST Project
WashU-HHMI Mouse EST Project
Washington University School of MedicineP
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:431424
Seq primer: -28ml3 rev2 ET from Amersham
High quality sequence stop: 499.
FEATURES
source
1..601
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone_lib="Barstead mouse pooled organs MPLRB4"
/sex="mixed"
/tissue_type="pooled organs"
/dev_stage="7 day"
/lab_host="DH10B"
/note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with
a modified polylinker; Site.1: EcoRI; Site.2: NotI; 1st
strand cDNA was primed with a Not I. oligo(dT) primer [5'
TGTTACAATCTGAGTGGAGCGCGCCCTTTTCTTTTTTTTTTTTTTTTT
3']; double-stranded cDNA was ligated to Eco RI adaptors
(GTGGATTCCGATCC), digested with Not I and cloned into
the Not I and Eco RI sites of the modified pT7T3 vector.
Library constructed by Bob Barstead."
BASE COUNT 129 a 177 c 163 g 132 t
ORIGIN

alignment_scores:
Quality: 917.50 Length: 183
Ratio: 5.126 Gaps: 2
Percent Similarity: 97.814 Percent Identity: 96.175

alignment_block:
US-09-462-625-2 x AA238752 ..
Align seg 1/1 to: AA238752 from: 1 to: 601

1 MetLeuPheAlaCysAlaLeuLeuAlaLeuGlyLeuAlaThrSerCy 17
|||||
55 ATGTTGTTTGGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTG 104
|||||
17 sSerPheLeuValProArgSerGluTrpArgAlaLeuProSerGluCysS 34
|||||
105 CAGTTTCATCTGTCGCCCGCAGTGTGAGTGGAGGCCCTCGCATCCGAGTGCT 154
|||||
34 eSerArgLeuGlyHisProValArgTyrValValIleSerHisPhrAla 50
|||||
155 CTAGCCGCTGGGGGAC. CCAGTTCGCTACGTGGTGTATCTCACACACAGCC 203
|||||
51 GlySerPheCysAsnSerProAspSerCysGluGlnAlaArgAsnVa 67
|||||
204 GGCACCTTCTGCACAGCCCGGACCTCTCTGTGNACAGCAGCGCCCGCAATGT 253
|||||
67 lGlnHisTyrHisLysAsnGluLeuGlyTrpCysAspValAlaTyrAsnP 84
|||||
254 GCAGCATTAACCAAGAATGAGCTGGGCTGGTGGATGTAGCTACCAACT 303
|||||
84 heLeuIleGlyGluAspGlyHisValTyrGluGlyArgGlyTrpAsnIle 100
|||||
304 TCTTATTGAGAGGAGCGGTGATGTATGAAAGCCGAGGCTGGAACATC 353
|||||
101 LysGlyAspHisThrGlyProIleTyrAsnProMetSerIleGlyIlePh 117
|||||
354 AAGGGTGACCACACAGCGGCCCATCTGGAATTCATGTCTATTGGCATCAC 403
|||||
117 rPheMetGlyAsnPheMetAspArgVal.ProAlaLysArgAlaLeuArg 133
|||||
404 CTTCTATGGGAACTTTCATGGACCGGTACCCG...CAAAGGGCCCTCCGT 450
|||||
134 AlaAlaLeuAsnLeuLeuGluCysGlyValSerArgGlyPheLeuArgSe 150
|||||
451 GCTGCCCTAAATCTCTCGAATGTGGGTGTCTCGGGGCTCTCTTGAGATC 500
|||||
150 rAsnTyrGluValLysGlyHisArgAspValGlnSerThrLeuSerProG 167
|||||
501 CAACTATGAAGTCCAAGAGCACCGGGATGTGCAAGCACTCTCTCTCCAG 550
|||||
167 lyAspGlnLeuTyrGlnValIleGlnSerTrpGluHisTyrArgGlu 182
|||||
551 GTGACAAACTCTATCAGTGCATCCAAAGCTGGGAACACTACCGAGAG 597
|||||
seq_name: gb_est1:AA238564
seq_documentation_block: 591 bp mRNA EST 03-MAR-1997
LOCUS AA238564
DEFINITION my35f04.r1 Barstead mouse pooled organs MPLRB4 Mus musculus cDNA
clone IMAGE:697855 5' similar to gb:X86374 M.musculus mRNA for TAG7
protein (MOUSE), mRNA sequence.
ACCESSION AA238564
VERSION AA238564.1 GI:1862588
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 591)
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.
TITLE The WashU-HHMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT Contact: Marra M/Mouse EST Project
WashU-HHMI Mouse EST Project
Washington University School of MedicineP
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:431415
Seq primer: -28ml3 rev2 ET from Amersham
High quality sequence stop: 499.
FEATURES
source
1..591
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone_lib="Barstead mouse pooled organs MPLRB4"
/sex="mixed"
/tissue_type="pooled organs"
/dev_stage="7 day"
/lab_host="DH10B"
/note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with
a modified polylinker; Site.1: EcoRI; Site.2: NotI; 1st
strand cDNA was primed with a Not I. oligo(dT) primer [5'
TGTTACAATCTGAGTGGAGCGCGCCCTTTTCTTTTTTTTTTTTTTTTT
3']; double-stranded cDNA was ligated to Eco RI adaptors
(GTGGATTCCGATCC), digested with Not I and cloned into
the Not I and Eco RI sites of the modified pT7T3 vector.
Library constructed by Bob Barstead."
BASE COUNT 129 a 177 c 163 g 132 t
ORIGIN

```

BASE COUNT	128 a	169 c	163 g	130 t	1 others
ORIGIN					

```

alignment_block:
US-09-462-625-2 x AA238564 ..
Align seq 1/1 to: AA238564 from: 1 to: 591

```

1 MetLeuPheAlaCysAlaLeuLeuAlaLeuLeuGlyLeuAlaThrSerCys 17
26 ATGTTGTTGGCTGCTGCTTCCCTGGCCCTCTGGGTCTGGCAACTCTCG 75
17 sserPheIleValProArgSerGluTrpArgAlaLeuProSerGluCysS 34
76 CAGTTTTCATCTGCGCCGCAAGTAGTGGAGGGGCCCTGGCA. TCCGAGTGT 124
34 erSerArgLeuGlyHisProValArgTyrValValIleSerHisThrAla 50
125 CTAGCGCGCTGGGAC. CCAGTTCGCTACGTGGTGATCTCACACAGCC 173
51 GlySerPheCysAsnSerProAspSerCysGluGlnGlnAlaArgAsnVa 67
174 GGCAGCTTCTGCAACAGCGCCGACTCTGTGACACAGACGCGCCGAATGT 223
67 lGlnHisTyrHisLysAsnGluLeuGlyTrpCysAspValAlaTyrAsnP 84
224 GCAGCATTTACCACAGAAATGAGCTGGGCTGGTGCATGTAGCCTACAACT 273
84 heLeuIleGlyGluAspGlyHisValTyrGluGlyArgGlyTrpAsnIle 100
274. TCTTATTGGAGAGACGGTCATGCTATGAAGCCGAGCGTGGAACTC 323
101 LysGlyAspHisThrGlyProIleTrpAsnProMetSerIleGlyIleTh 117
324 AAGGGTGAACACACAGGGCCCATCTGGAAATCCCATGCTATTGGCATCAC 373
117 rPheMetGlyAsnPheMetAspArgValProAlaLysArgAlaLeuArg. 133
374 CTTTCATGGGGAACCTTCATGGACCGGTATCCGCCAAGGGTGCCTCCGTG 422
134 AlaAlaLeuAsnLeuLeuGluCysGlyValSerArgGlyPheLeuArgSe 150
424 GCTGCCCTAAATCTCTTGGAAATGGGGGTCTCTGGGGCTTCTTTGAGATC 473
150 rAsnTyrGluValLysGlyHisArgAspValGlnSerThrLeuSerProg 167
474 CAACTATGAAGTCAAGGACACCGGGATGTGCAAGCATCTTCTCTCCAG 522
167 lyAspGlnLeuTyrGlnValIleGlnSerTrpGluHisTyrArgGlu 182
524 GTGACCAACTCTATCAGTCTCATCCAAAGCTGGGAACACTACCGAGAG 570

Result No.	Query			DB	ID	Description
	Score	Match	Length			
1	201	20.2	539	2	G70520	probable csp prote
2	130	13.1	308	2	T35303	hypothetical prote
3	119	11.9	151	1	MUBPA7	N-acetyl-muramoyl-L
4	118.5	11.9	151	2	S07506	N-acetyl-muramoyl-L
5	104	10.4	44	2	S54144	tmsh protein - mou
6	84	8.4	702	2	G83443	hypothetical prote
7	79	7.9	445	2	C89233	lysyl endopeptidas
8	78.5	7.9	1376	2	G00043	osteonidogen - hum
9	78	7.8	424	2	T26751	hypothetical prote
10	77.5	7.8	699	2	T33375	hypothetical prote
11	77	7.7	526	2	S34945	nitrogenase (EC 1.
12	75.5	7.6	726	2	S61689	hypothetical prote
13	74.5	7.5	234	2	S44106	arginine-specific
14	74.5	7.5	276	2	T09041	ciml protein homol
15	74.5	7.5	991	2	I40229	arginyl endopeptid
16	74.5	7.5	1526	2	S4763	gangipain R (EC 3.
17	74.5	7.5	1704	2	A54426	gangipain R (EC 3.
18	74.5	7.5	3535	2	E83641	probable hemagglut
19	74	7.4	416	2	S76310	hypothetical prote
20	74	7.4	863	2	S74447	ferrichrome-iron r
21	73.5	7.4	188	2	E83080	beta-lactamase exp
22	73.5	7.4	3421	1	WZBE66	367k tegument prot
23	73	7.3	298	2	T72522	hypothetical prote
24	72.5	7.3	242	2	T34767	hypothetical prote
25	72.5	7.3	356	2	D65096	hypothetical 40.1
26	72.5	7.3	356	2	H85968	hypothetical prote
27	72.5	7.3	656	2	D82388	glycogen operon pr
28	72	7.2	260	1	VCBW	coat protein - sou
29	72	7.2	405	2	A83038	hypothetical prote

C; Keywords: hydrolase

Query Match 11.9%; Score 119; DB 1; Length 151;
Best Local Similarity 33.7%; Pred. No. 0.00024;
Matches 33; Conservative 18; Mismatches 37; Indels 10; Gaps 4;

QY 65 RNVOHYHKNELGWCDAVYNFLIGEDGHVYEGRGNKGDHTGPINWPMISGITFTMGFMFD 124
::: |||: || | | : | | | | | | | | | | | | | | | | | | | | | |
Ddb 31 REIRQWHKDQ GWLWDVGYPHFIKRDTVEAGRDEMVAVGSHAKG YNHNSIGVCLVGGIDD 88
::: |||: || | | : | | | | | | | | | | | | | | | | | | | | | |

QY 125 R-----VPAAK-RALRAALNLECGVSRGFLRSNYEV 154
|||: |||: || | | : | | | | | | | | | | | | | | | | | | | | | |
Ddb 89 KGKFEDANFTPAQMQLSRSLTLTLLAKYEGAGLRHAHEV 126
|||: |||: || | | : | | | | | | | | | | | | | | | | | | | | | |

RESULT 4

S07506 N-acetylmuramoyl-L-alanine amidase (EC 3.5.1.28) - phase T3
N:Alternate names: gene 3.5 protein
C:Species: phage T3
C>Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 22-Jun-1999
C:Accession: S07506
R:Beck, P.J.; Gonzalez, S.; Ward, C.L.; Molineux, I.J.
J. Mol. Biol. 210, 687-701, 1989
A:Title: Sequence of bacteriophage T3 DNA from gene 2.5 through gene 9.
A:Reference number: S07500; MUID:90133923
A:Accession: S07506
A:Molecule type: DNA
A:Residues: 1-151 <BC>
A:Cross-references: EMBL:X17255; NID:g15682; PIDN:CAA35133.1; PID:g15695
C:Genetics:
A:Gene: 3.5
C:Superfamily: phage T7 N-acetylmuramoyl-L-alanine amidase
C:Keywords: hydrolase

Query Match 11.9%; Score 118.5; DB 2; Length 151;
Best Local Similarity 30.4%; Pred. No. 0.00027;
Matches 38; Conservative 19; Mismatches 43; Indels 25; Gaps 7;

QY 65 RNVOHYHKNELGWCDAVYNFLIGEDGHVYEGRGNKGDHTGPINWPMISGITFTMGFMFD 124
::: |||: || | | : | | | | | | | | | | | | | | | | | | | | | |
Ddb 31 REIRQWHK-EQWLWDVGYPHFIKRDTVEAGRDELAVGSHAKG-YNHNSIGVCLVGGIDD 88
::: |||: || | | : | | | | | | | | | | | | | | | | | | | | | |

QY 125 R-----VPAAK-RALRAALNLECGVSRGFLRSNYE---VKHRDVQSTLSPGDQLYQV 173
|||: |||: || | | : | | | | | | | | | | | | | | | | | | | | | |
Ddb 89 KGKFEDANFTPAQMQLSRSLLT-----LLAKYEGSVLRHHADVAPKAPS---FDL 136
|||: |||: || | | : | | | | | | | | | | | | | | | | | | | | | |

QY 174 IQSWE 178
::: |||
Ddb 137 KRWE 141
::: |||

RESULT 5

S54144 tmsl protein - mouse
C:Species: Mus musculus (house mouse)
C>Date: 08-Jul-1995 #sequence_revision 21-Jul-1995 #text_change 21-Jul-1995
C:Accession: S54144
R:Kustikova, O.S.
submitted to the EMBL Data Library, April 1995
A:Reference number: S54144
A:Accession: S54144
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-44 <KUS>
A:Cross-references: EMBL:X86374

Query Match 10.4%; Score 104; DB 2; Length 44;
Best Local Similarity 95.5%; Pred. No. 0.0016;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MLFACALLALLGLATSCSFIVP 22
|||||
Db 1 MLFACALLALLGLATSCSFICP 22

RESULT 6

G83443
hypothetical protein PA1613 [imported] - Pseudomonas aeruginosa (strain PA01)
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: G83443
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; B.
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
A:Reference number: A82950; MUID:20437337
A:Accession: G83443
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-702 <STO>
A:Cross-references: GB:AE004589; GB:AE004091; NID:g9947574; PIDN:AAG05002.1; GSPDB:GN001
A:Experimental source: strain PA01
C:Genetics:
A:Gene: PA1613

Query Match 8.4%; Score 84; DB 2; Length 702;
Best Local Similarity 24.0%; Pred. No. 3.5; Indels 56; Gaps 13;
Matches 43; Conservative 24; Mismatches 56; Indels 56; Gaps 13;

QY 25 EWRLPSECSRLGHPVRYVVISHTAGSCFNSPDCSCQQAQVQHYHKNELGW---CDVA 81
|||
Db 242 EWQLAAKESNQ---HPVAY-----RSPFPQVDDVD-NSDTLAHDYAGSLRNVDLDP 292
|||
QY 82 YNPLIGEDGHVYEGRGWNIKDHGTPIWNPMSIGITP-----MGNFMD--- 124
|||
Db 293 HSPYI---QGY---AQHWERRQE-----WNACEAGIAFSPQLAQLWELNPDYLRFFADDDPT 342
|||
QY 125 RVP--AKRALRALNLE-----CG-VSRGFLRSNVEYKGRDVGOSTLSPGDOL 170
|||
Db 343 RLPPGSPQELQATQVLDQLRNSGLRVPVCGVRDNRINRESYDL-----EMQDTLSLGDNL 397
|||

RESULT 7

C69233
lysyl endopeptidase - Methanobacterium thermoautotrophicum (strain Delta H)
C:Species: Methanobacterium thermoautotrophicum
C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 22-Oct-1999
C:Accession: C69233
R:Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldredge, T.;
.; Qiu, D.; Spadafora, R.; Vicaire, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiwani, N.
ki, S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.
J. Bacteriol. 179, 7135-7155, 1997
A:Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: funct
A:Reference number: A69000; MUID:98037514
A:Accession: C69233
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-445 <MTH>
A:Cross-references: GB:AE000872; GB:AE000666; NID:g2622082; PIDN:AA85492.1; PID:g262209
A:Experimental source: strain Delta H
C:Genetics:
A:Gene: MTH995
A:Start codon: GTG

Query Match 7.9%; Score 79; DB 2; Length 445;
Best Local Similarity 26.3%; Pred. No. 6.3; Indels 40; Gaps 7;
Matches 40; Conservative 17; Mismatches 55; Indels 40; Gaps 7;

QY 34 SSRGLGHPVRYVVISHTAGSCFNSPDCSCQQAQVQHYHKNELGWCDVAYNFI-- 86

Db 17 ASVITFPVAEAAHILIVADSLSDSPMYD-EAKKVATTLKNRGYHVLELYNNATVKNILK 75
|||
QY 87 ---GEDGHVYEGRGWNIKDHGTGP-----IWNP-----MSIGITFMGNFM 123
|||
Db 76 GMYGADAVIYAGGGYAGHYDGRGNATPPFGIVARDGYIWSAGDMMVNGITFYAPFK 135
|||
QY 124 DRVPAKRALRAALNLLLECGVSRGFLRSNVEVK 155
|||
Db 136 DGIPA-----ILLHACFTSGWV-EYEYK 158
|||

RESULT 8

G00043
osteonidogen - human
C:Species: Homo sapiens (man)
C:Date: 13-Mar-1997 #sequence_revision 13-Mar-1997 #text_change 05-Nov-1999
C:Accession: G00043
R:Ohno, I.
submitted to GenBank, July 1996
A:Reference number: H00043
A:Accession: G00043
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1376 <OHN>
A:Cross-references: GB:D86425; NID:gl49166; PIDN:BAAL13087.1; PID:gl449167
C:Superfamily: nidogen; EGF homology; LDL receptor YWTD-containing repeat homology; t
F:806-843/Domain: EGF homology <EGF1>
F:853-891/Domain: EGF homology <EGF>
F:941-1006/Domain: thyroglobulin type I repeat homology <THY1>
F:1020-1085/Domain: thyroglobulin type I repeat homology <THY2>
F:1242-1286/Domain: LDL receptor YWTD-containing repeat homology <YW3>

Query Match 7.9%; Score 78.5; DB 2; Length 1376;
Best Local Similarity 25.2%; Pred. No. 27;
Matches 38; Conservative 10; Mismatches 46; Indels 57; Gaps 9;

QY 3 FACALL-ALLGLATSCSFIVPRSEWRALPSECSRLGHPVRYVVISHTAGSF---CNS-- 56
|||
Db 876 FSCACLPYAGDGHQCTDV-----DECSENRCHPA--ATCYNTPGFSFCRCOPGY 923
|||
QY 57 -----PDS-----CEQAQNVQHYH-----KNELGWCDVAYNFI----- 85
|||
Db 924 YGDGFQICIPDSTSLTPECQQAQYAYPGARFIPQCDEQGNFLPLQCHGSTGFCW 983
|||
QY 86 -IGEDGHVYEGRGWNIKDHGTPIWNPMSIG 115
|||
Db 984 CVDPDGH-----EVPQTQTPPGSTPPHCG 1007
|||

RESULT 9

T26751
hypothetical protein Y39A1C.4 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T26751
R:Wall, M.
submitted to the EMBL Data Library, June 1998
A:Reference number: Z20259
A:Accession: T26751
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-424 <WIL>
A:Cross-references: EMBL:AL023839; PIDN:CAA19506.1; GSPDB:GN00021; CESP:Y39A1C.4
A:Experimental source: clone Y39A1C
C:Genetics:
A:Gene: CESP:Y39A1C.4
A:Map position: 3
A:Introns: 65/3; 203/3; 320/3; 372/1

Query Match 7.8%; Score 78; DB 2; Length 424;

DB 645 YGCVTLHIECLIGKDIYMKPGSSWLFKGRKRVLRNNHRTPLCRCKCKDCPCPKALVPRK 704

QY 134 RAINLLECGVSR 145
: | | | |
DB 705 STLNFCSTGCMR 716

RESULT 13

S41406

arginine-specific cysteine proteinase - Porphyromonas gingivalis

C:Species: Porphyromonas gingivalis

C:Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 08-Oct-1999

C:Accession: S41406

R:Stotiropoulos, C.; Kirszbaum, L.; Slakeski, N.; Jackson, C.; Reynolds, E.
submitted to the EMBL Data Library, January 1994

A:Description: Complete nucleotide sequence of the prtR gene encoding an arginine-spe

A:Reference number: S41406

A:Accession: S41406

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-234 <SOT>

A:Cross-references: EMBL:Z29570; NID:g450920; PIDN:CAA82674.1; PID:g450921

Query Match 7.5%; Score 74.5; DB 2; Length 234;
Best Local Similarity 26.1%; Pred. No. 8.1;
Matches 23; Conservative 15; Mismatches 27; Indels 23; Gaps 4;

QY 118 FMGNFMDRVPK--RALRAALNLE-----CGVSRGFLRSNYEVKG-----H 157
DB 29 YEGDIKDFVDWKNQRLRTEVKVAEDIASPVTANAIOQFVKQYEKEGNDLTYVLLVGDH 88
QY 158 RDVQSTLSPG---DQLYQVIOQWEHYRE 182
DB 89 KDIPAKITPGIKSDQVYGIVGNDHYNE 116

RESULT 14
T09041
c1ml protein homolog F26K10.130 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 17-Nov-2000
C:Accession: T09041
R:Bevan, M.; Lennard, N.; Quail, M.; Harris, B.; Rajandream, M.A.; Barrell, B.G.; Bancroft, submitted to the Protein Sequence Database, June 1999
A:Reference number: Z16533
A:Accession: T09041
A:Molecule type: DNA
A:Residues: 1-276 <REV>
A:Cross-references: EMBL:AL049803; GSPDB:GNO0062; ATSP:F26K10.130
A:Experimental source: cultivar Columbia; BAC clone F26K10
C:Genetics:
A:Gene: ATSP:F26K10.130
A:Map position: 4
A:Introns: 54/1; 154/2; 197/3; 235/1
C:Superfamily: expansin

Query Match 7.5%; Score 74.5; DB 2; Length 276;
Best Local Similarity 27.6%; Pred. No. 9.8;
Matches 32; Conservative 15; Mismatches 28; Indels 41; Gaps 7;
QY 59 SCEQOARVQHYHKNELGWCDAVNFILIG-EDG-----HYVEGR 96
DB 156 ACKYRGKNIA-FHYNE-GSTDFWLSLVEFEGEDIGSMHIRQAGAREWLEMKHYW-GA 212
QY 97 GWNITKDHGTPIWNPMSIGITFMGNFMDRV-----PAKRALRAALNLL 139
DB 213 NWCIIIG---GPLKGPFSIKLTLTAERNRVTAGRRMETVRKVPNQRLRLSLFL 265

RESULT 15
I40229
arginyl endopeptidase - Porphyromonas gingivalis
C:Species: Porphyromonas gingivalis
C:Date: 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change 08-Oct-1999
C:Accession: I40229
R:Okamoto, K.; Misumi, Y.; Kadowaki, T.; Yoneda, M.; Yamamoto, K.; Ikehara, Y.
Arch. Biochem. Biophys. 316, 917-925, 1995
A:Title: Structural characterization of arginylpain, a novel arginine-specific cysteine
A:Reference number: I40229; MUID:9516884
A:Accession: I40229
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-991 <RES>
A:Cross-references: GB:D26470; NID:g927644; PIDN:BAA05484.1; PID:g927645

Query Match 7.5%; Score 74.5; DB 2; Length 991;
Best Local Similarity 26.1%; Pred. No. 44;
Matches 23; Conservative 15; Mismatches 27; Indels 23; Gaps 4;

QY 118 FMGNFMDRVPK--RALRAALNLE-----CGVSRGFLRSNYEVKG-----H 157
DB 247 YEGDIKDFVDWKNQRLRTEVKVAEDIASPVTANAIOQFVKQYEKEGNDLTYVLLVGDH 306

QY 158 RDVQSTLSPG---DQLYQVIOQWEHYRE 182
DB 307 KDIPAKITPGIKSDQVYGIVGNDHYNE 334
Search completed: December 17, 2001, 07:49:02
Job time: 25 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 17, 2001, 07:48:37 ; Search time 11.77 Seconds
(without alignments)
566.950 Million cell updates/sec

Title: US-09-462-625-2

Perfect score: 996

Sequence: 1 MLFACALLALLGLATSCSFI.....TLSPGDQLYQVDSWEHYRE 182

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	121	12.1	150	1 NAAA_BPT7	P00806 bacterioph
2	118.5	11.9	150	1 NAAA_BPT3	P20331 bacterioph
3	78.5	7.9	1375	1 NID2_HUMAN	Q14112 homo sapien
4	77	7.7	526	1 ANFD_RHOCA	Q07933 rhodobacter
5	76.5	7.7	197	1 GSHL_PIG	P36968 sus scrofa
6	74.5	7.5	736	1 CPG2_PORGI	P95493 porphyromon
7	74.5	7.5	991	1 CPGL_PORGI	P28784 porphyromon
8	73.5	7.4	3421	1 TEGU_HSVB	P28955 equine heip
9	72.5	7.3	356	1 YGJU_ECOLI	P42591 escherichia
10	72	7.2	279	1 COAT_SBMV	P03607 southern be
11	72	7.2	280	1 ANFD_HELGE	O68955 heliobacter
12	72	7.2	395	1 ASAH_HUMAN	Q13510 homo sapien
13	71	7.1	826	1 LPN2_MYXXA	P36774 myxococcus
14	71	7.1	832	1 DPOL_HRPVY	P03156 hepatitis b
15	70	7.0	478	1 NIFE_METTH	O27607 methanobact
16	70	7.0	483	1 NIFE_METMP	P71528 methanococc
17	70	7.0	518	1 ANFD_AZQVI	P16266 azotobacter
18	70	7.0	669	1 AMY_ALTHA	P29957 alteromonas
19	69.5	7.0	157	1 GSHH_MOUSE	O70325 mus musculu
20	69.5	7.0	480	1 LSTP_STAST	P10548 staphylococ
21	69.5	7.0	493	1 LSTP_STASI	P10547 staphylococ
22	69	6.9	426	1 WECF_SALTY	P37458 salmonella
23	69	6.9	509	1 CP41_RAT	P08516 rattus norv
24	68.5	6.9	374	1 Y483_TREPA	O83496 treponema p
25	68.5	6.9	389	1 ALC_XENLA	Q9W6S5 xenopus lae
26	68	6.8	488	1 GDNH_RAT	O62997 rattus norv
27	68	6.8	5376	1 ZAN_MOUSE	O88799 mus musculu
28	67.5	6.8	197	1 GSHH_HUMAN	P36969 homo sapien
29	67.5	6.8	524	1 CHID_BACCI	P27050 bacillus ci
30	67.5	6.8	579	1 CCBS_DAUCA	Q04647 daucus caro
31	67	6.7	428	1 WN8B_XENLA	P31291 xenopus lae
32	67	6.7	484	1 GLGA_BACSU	P39125 bacillus su
33	67	6.7	630	1 YND1_YEAST	P40009 saccharomyc

34	66.5	6.7	1429	1 NOS1_RAT	P29476 rattus norv
35	66.5	6.7	1534	1 MTDM_ARATH	P34881 arabidopsis
36	66.5	6.7	3083	1 POLG_ZYMYR	Q89330 z genome po
37	66	6.6	270	1 Alys_BPRIT	Q38135 bacterioph
38	66	6.6	347	1 VCAA_BPT3	P19693 bacterioph
39	66	6.6	376	1 YB1L_YEAST	P38287 saccharomyc
40	66	6.6	433	1 VCAB_BPT3	P19728 bacterioph
41	66	6.6	530	1 YDEV_ECOLI	P77432 escherichia
42	66	6.6	711	1 CDGT_BACST	P31797 bacillus st
43	66	6.6	773	1 HEXB_ALTSO	P49007 alteromonas
44	66	6.6	777	1 LON_BUCAL	P57549 buchnera ap
45	66	6.6	832	1 DFOM_HRPVY	P04484 hepatitis b

ALIGNMENTS

RESULT 1					
NAAA_BPT7					
ID	NAAA_BPT7	STANDARD;	PRT;	150 AA.	
AC	P00806; Q38567;				
DT	21-JUL-1986 (Rel. 01, Created)				
DT	01-FEB-1995 (Rel. 31, Last sequence update)				
DT	20-AUG-2001 (Rel. 40, Last annotation update)				
DE	N-ACETYLURAMUOYL-L-ALANINE AMIDASE (EC 3.5.1.28) (T7 LYSOZYME).				
GN	3.5.				
OS	Bacteriophage T7.				
OC	Viruses; dsDNA viruses, no RNA stage; Tailed phages; Podoviridae;				
OC	T7-like phages.				
OX	NCBI_TaxID=10760;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=83241725; PubMed=6864790;				
RA	Dunn J.J., Studier F.W.;				
RT	"Complete nucleotide sequence of bacteriophage T7 DNA and the				
RT	locations of T7 genetic elements."				
RL	J. Mol. Biol. 166:477-535(1983).				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=82078034; PubMed=7310871;				
RA	Dunn J.J., Studier F.W.;				
RT	"Nucleotide sequence from the genetic left end of bacteriophage T7				
RT	DNA to the beginning of gene 4."				
RL	J. Mol. Biol. 148:303-330(1981).				
RN	[3]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=95099869; PubMed=7801634;				
RA	Huang W., Cui X., Tian Y., Lin M., Peng X.;				
RT	"Cloning of T7 lysozyme gene and construction of the vector for				
RT	transgenic plants resistant to bacterial infection."				
RL	Wei Sheng Wu Hsueh Pao 34:261-265(1994).				
RN	[4]				
RP	X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS), AND REVISION TO 118.				
RX	MEDLINE=94224877; PubMed=8171031;				
RA	Cheng X., Zhang X., Pflugrath J.W., Studier F.W.;				
RT	"The structure of bacteriophage T7 lysozyme, a zinc amidase and an				
RT	inhibitor of T7 RNA polymerase."				
RL	Proc. Natl. Acad. Sci. U.S.A. 91:4034-4038(1994).				
RN	[5]				
RP	X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS) OF COMPLEX WITH POLYMERASE.				
RX	MEDLINE=98336199; PubMed=9670025;				
RA	Jeruzalmi D., Steitz T.A.;				
RT	"Structure of T7 RNA polymerase complexed to the transcriptional				
RT	inhibitor T7 lysozyme."				
RL	EMBO J. 17:4101-4113(1998).				
CC	-!- FUNCTION: THIS PROTEIN IS NOT REQUIRED FOR LYSIS BUT PLAYS AN				
CC	IMPORTANT ROLE IN DNA SYNTHESIS. IT IS KNOWN TO DETACH THE HOST				
CC	CHROMOSOME FROM THE BACTERIAL MEMBRANE TO WHICH IT IS NORMALLY				
CC	BOUND. IT IS A BIFUNCTIONAL PROTEIN THAT CUTS AMIDE BONDS IN THE				
CC	BACTERIAL CELL WALL AND BINDS TO AND INHIBITS TRANSCRIPTION BY				
CC	T7 RNA POLYMERASE.				
CC	-!- CATALYTIC ACTIVITY: HYDROLYZES THE LINK BETWEEN N-ACETYLURAMUOYL				
CC	RESIDUES AND L-AMINO ACID RESIDUES IN CERTAIN BACTERIAL CELL-WALL				

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
at the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announcements>)
or send an email to license@isb-sib.ch).

[illegible]

FT	57	59
STRAND		
TURN	63	64
FT	65	65
STRAND		
TURN	71	72
FT	74	76
HELIX		
FT	77	82
STRAND		
FT	85	86
STRAND		
TURN	88	89
FT	92	92
STRAND		

```

SQ SEQUENCE      150 AA;   16948 MW;    6E4F5D183C9EEB5Z CRC64;

Query Match          12.1%; Score 121; DB 1; Length 150;
Best Local Similarity 34.7%; Pred. No. 5.8e-05;
Matches 34; Conservative 16; Mismatches 38; Indels 10; Gaps 1

QY 65 RNVOHYHKNELGWCDAVNFLLIGEDGHVYEGRGNRIKGDHTGTPIWNPMSGITTFMGNFMFD 1
```

```

Db      30  RETRQWHK  EQGWLVDGYHF1IKRDGTVACGRDEMAVGSHAKG-YNHNSIGVCLVGGIDD
      125  R-----VPAK-RALRAALNLECGVSRGFLRSNYEV 154
      :      |||  :|||  |  |  |||:||||
Db      88  KGKFDANFTPAQMQSLRSLVTLAKYEGAVLRAMEV 125

```

CC -1- ENZYME REGULATION: BINDING OF 3' NUCLEOTIDE
CC ACTIVITY (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE N-ACETYLMURAMYL-L-ALANINE AMIDASE
CC FAMILY 2.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC derived from, or based on, the removed sequence and for commercial
CC use by and for commercial

	Query Match	11.9%	Score 118.5;	DB 1;	Length 150;
	Best Local Similarity	30.4%;	Pred. No. 0.0001;		
	Matches	38;	Conservative	19;	Mismatches 43; Indels 25; Gaps 7;
QY	65	RNVQHYYKKNELGWCDDVAYNFLIGEDGHVYEGRGWNKIDGTGPTWNPMSIGITPWGNFMD	124		
	30	PRTPBQHK-FCQGLNDVGXHFILIKRDTGYEAGDELAVGSHAKG-YNNHNSIGVLGVGGIDD	87		

125 R-----VPAK-RAIRAAINLECGVSRGFLRSNYE---VKGHRDVQSTLSPGDQLYQV 173
 : ||: :||: | | | | : : | | | : : :
 88 KGKFDANFTPAQMOSLRSLLVT-----LLAYEGSVLRAHDVAPKACPS---FDL 135

QY 174 IOSWE 178
DB 136 KRWE 140

RESULT 3

```

NID2_HUMAN
ID Q14112; O43710; STANDARD; PRT; 1375 AA.
DT 15-DEC-1998 (Rel. 37, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DE NID2_HUMAN (Rel. 40, Last annotation update)
DE NID2_HUMAN-2 PRECURSOR (NID-2) (OSTEONIDOGEN).
GN NID2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF N-TERMINUS.
RX MEDLINE=98406162; PubMed=9733643;
RA Kohfeldt E., Sasaki T., Goehring W., Timpl R.;
RT "Nidogen-2: a new basement membrane protein with diverse binding
RT properties."
RL J. Mol. Biol. 282:99-109(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Cancellous bone;
RA Ohno I., Hashimoto J., Takaoka K., Ochi T., Okubo K., Matsubara K.;
RT "The cloning and characterization of a cDNA for the novel bone matrix
RT protein: osteonidogen."
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Ohno I., Okubo K., Matsubara K.;
RT "Human osteonidogen gene: intron-exon junctions and chromosomal
RT localization."
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: CELL ADHESION GLYCOPROTEIN WHICH IS WIDELY DISTRIBUTED
CC IN BASEMENT MEMBRANES. BINDS TO COLLAGENS I AND IV. TO PERLECAN
CC AND TO LAMININ 1. DOES NOT BIND FIBRILINS. IT PROBABLY HAS A ROLE
CC IN CELL-EXTRACELLULAR MATRIX INTERACTIONS.
CC -!- TISSUE SPECIFICITY: HEART, PLACENTA AND BONE. LESS IN PANCREAS,
CC KIDNEY AND SKELETAL MUSCLE.
CC -!- PTM: HIGHLY N- AND O-GLYCOSYLATED.
CC -!- SIMILARITY: CONTAINS 5 EGF-LIKE DOMAINS.
CC -!- SIMILARITY: CONTAINS 2 THYROGLOBULIN TYPE-I DOMAINS.
CC -!- SIMILARITY: CONTAINS 5 LDL-RECEPTOR YWTD DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AJ223350; CA111418.1;
DR EMBL; D86425; BAA13087.1;
DR EMBL; AB009799; BAA24112.1;
DR EMBL; AB009778; BAA24112.1; JOINED.
DR EMBL; AB009779; BAA24112.1; JOINED.
DR EMBL; AB009780; BAA24112.1; JOINED.
DR EMBL; AB009781; BAA24112.1; JOINED.
DR EMBL; AB009782; BAA24112.1; JOINED.
DR EMBL; AB009783; BAA24112.1; JOINED.
DR EMBL; AB009784; BAA24112.1; JOINED.
DR EMBL; AB009785; BAA24112.1; JOINED.
DR EMBL; AB009786; BAA24112.1; JOINED.
DR EMBL; AB009787; BAA24112.1; JOINED.
DR EMBL; AB009788; BAA24112.1; JOINED.
DR EMBL; AB009789; BAA24112.1; JOINED.

```

```

DR EMBL; AB009790; BAA24112.1; JOINED.
DR EMBL; AB009791; BAA24112.1; JOINED.
DR EMBL; AB009792; BAA24112.1; JOINED.
DR EMBL; AB009793; BAA24112.1; JOINED.
DR EMBL; AB009794; BAA24112.1; JOINED.
DR EMBL; AB009795; BAA24112.1; JOINED.
DR EMBL; AB009796; BAA24112.1; JOINED.
DR EMBL; AB009797; BAA24112.1; JOINED.
DR EMBL; AB009798; BAA24112.1; JOINED.
DR MIM; 605399;
DR InterPro; IPR000152; Asx_hydroxyl.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR000033; Ldl_receptor_rep.
DR InterPro; IPR003886; Nidogen_ext.
DR InterPro; IPR000716; Thyroglobulin_1.
DR Pfam; PF00008; EGF; 5.
DR Pfam; PF00058; ldl_recept_b; 4.
DR Pfam; PF00086; thyroglobulin_1; 2.
DR SMART; SM00179; EGF_CA; 2.
DR SMART; SM00001; EGF-like; 3.
DR SMART; SM01135; LY; 4.
DR SMART; SM00539; NIDO; 1.
DR SMART; SM00211; TY; 2.
DR PROSITE; PS00010; ASX-HYDROXYL; 3.
DR PROSITE; PS00484; THYROGLOBULIN_1; 2.
DR PROSITE; PS01186; EGF-2; 4.
DR PROSITE; PS01187; EGF_CA; 2.
KW Basement membrane; Extracellular matrix; Glycoprotein; Signal;
KW Calcium-binding; Repeat; EGF-like domain; Cell adhesion.
FT SIGNAL 1
FT CHAIN 31 1375
FT DOMAIN 484 524 NIDOGEN-2.
FT DOMAIN 759 800 EGF-LIKE 1.
FT DOMAIN 801 843 EGF-LIKE 2.
FT DOMAIN 848 891 EGF-LIKE 3.
FT DOMAIN 892 930 EGF-LIKE 4.
FT DOMAIN 936 1006 EGF-LIKE 5.
FT DOMAIN 1015 1085 THYROGLOBULIN TYPE I 1.
FT DOMAIN 1157 1200 THYROGLOBULIN TYPE I 2.
FT DOMAIN 1201 1243 LDL-RECEPTOR YWTD MOTIF 1.
FT DOMAIN 1244 1288 LDL-RECEPTOR YWTD MOTIF 2.
FT DOMAIN 1289 1328 LDL-RECEPTOR YWTD MOTIF 3.
FT DOMAIN 1329 1373 LDL-RECEPTOR YWTD MOTIF 4.
FT DOMAIN 1376 1776 LDL-RECEPTOR YWTD MOTIF 5.
FT DISULFID 763 776 BY SIMILARITY.
FT DISULFID 770 786 BY SIMILARITY.
FT DISULFID 788 799 BY SIMILARITY.
FT DISULFID 805 818 BY SIMILARITY.
FT DISULFID 812 827 BY SIMILARITY.
FT DISULFID 829 842 BY SIMILARITY.
FT DISULFID 852 867 BY SIMILARITY.
FT DISULFID 859 877 BY SIMILARITY.
FT DISULFID 879 890 BY SIMILARITY.
FT DISULFID 896 907 BY SIMILARITY.
FT DISULFID 901 916 BY SIMILARITY.
FT DISULFID 918 929 BY SIMILARITY.
FT CARBOHYD 417 417 N-LINKED (GLCNAC. . .) (PROBABLE).
FT CARBOHYD 658 658 N-LINKED (GLCNAC. . .) (PROBABLE).
FT CARBOHYD 693 693 N-LINKED (GLCNAC. . .) (PROBABLE).
FT CARBOHYD 703 703 N-LINKED (GLCNAC. . .) (PROBABLE).
FT CARBOHYD 1124 1124 N-LINKED (GLCNAC. . .) (PROBABLE).
FT CONFLICT 44 44 G -> W (IN REF. 3).
FT CONFLICT 54 172 ESSAYVKLANPLHFYEAFNSLVGNTGIISTODFPRETOY
VYDFPTDFPAIPAFADIDITSHGRGVLYREDTSAPVLGL
AARYVRAGEFSPARFTTHAFATWEQVGAEEVKRG ->
VKLSRGEAGESPALTKPDSATSWAPTASPLSTSEKRS
MTMISPTSRPSPLEWSTRTATAEACTERTPPQCCWA
WPPAMCALASRALRAFYPHRPLFCHLGAGRLRGQTR
(IN REF. 2 AND 3).
D -> G (IN REF. 1).
FT CONFLICT 453 453
FT SEQUENCE 1375 AA; 151394 MW; 414299D244205FBC CRC64;
SQ

```

QY 54 CNSPSCSEQOARVOHYHKKNELGWCVDAYNFL-----IGEDGHVYEGRGNWIKGDHTGP I - 108
||||| :||| :||| :||| :||| :
Db 166 CNSPGFAGPSQS- GHKINTAWLNOKVGVEPDYLGEHVINTVGE-YNIOGDQEWID 222

QY 109 -WNPMISIGI--TFMGFNMDRVPAKRAL-RAAALNFLECGVSRGF-----LRSNYEV 154
:||| :||| :||| :||| :||| :
Db 223 YFRNGIQVLSTFTGN--GSYDLSLRMMHRAHLNVLECARSAEYCDELARYGI 274

RESULT 5

GSHH_PIG STANDARD; PRT; 197 AA.

ID GSHH_PIG AC P36968;
DT 01-JUN-1994 (Rel. 29, Created)
DT DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT DE 20-AUG-2001 (Rel. 40, Last annotation update)
DE PHOSPHOLIPID HYDROPEROXIDE GLUTATHIONE PEROXIDASE, MITOCHONDRIAL
DE PRECURSOR (EC 1.11.1.9) (PGPX) (GPX-4).
GN GPX4.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;

[1]
SEQUENCE FROM N.A.
MEDLINE=93312346; PubMed=8323565;
RA Sundin R.A., Dyer J.A., Moran T., Evenson J.K., Sugimoto M.;
RX "phospholipid hydroperoxide glutathione peroxidase: full-length pig
blastocyst cDNA sequence and regulation by selenium status.";
RL Biochem. Biophys. Res. Commun. 193:905-911(1993).

[2]
SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
TISSUE=Liver, and Heart;
RX MEDLINE=9411752; PubMed=8125951;
RA Brigelius-Flohe R., Aumann K.-D., Bloeker H., Gross G., Ursini F.,
RA Kloepfel K.-D., Maiorino M., Roveri A., Schuckelt R., Ursini F.,
RA Wingender E., Flohe L.;
RX "Phospholipid-hydroperoxide glutathione peroxidase. Genomic DNA,
cDNA, and deduced amino acid sequence.";
RL J. Biol. Chem. 269:7342-7348(1994).

[3]
SEQUENCE OF 40-197 FROM N.A., AND PARTIAL SEQUENCE.
TISSUE=Heart;
RX MEDLINE=92137773; PubMed=1778506;
RA Strassburger W., Ursini F., Wolf B., Flohe L.;
RA "phospholipid hydroperoxide glutathione peroxidase is a selenoenzyme
distinct from the classical glutathione peroxidase as evident from
cDNA and amino acid sequencing.";
RL Free Radic. Res. Commun. 14:343-361(1991).
CC -1- FUNCTION: COULD PLAY A MAJOR ROLE IN PROTECTING MAMMALS FROM THE
TOXICITY OF INGESTED LIPID HYDROPEROXIDES.
CC -1- CATALYTIC ACTIVITY: 2 GLUTATHIONE + H(2)O(2) = OXIDIZED
GLUTATHIONE + 2 H(2)O.
CC -1- COFACTOR: SELENOCYSTEINE. THE ACTIVE-SITE SELENOCYSTEINE IS
ENCODED BY THE OPAL CODON, UGA.
CC -1- SUBUNIT: MONOMER.
CC -1- SUBCELLULAR LOCATION: MITOCHONDRIAL AND CYTOSOLASMIC.
CC -1- ALTERNATIVE PRODUCTS: A SINGLE NUCLEAR GENE PRODUCES BOTH FORMS
BY USE OF ALTERNATIVE INITIATION CODONS IN THE SAME READING FRAME.
CC -1- SIMILARITY: BELONGS TO THE GLUTATHIONE PEROXIDASE FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch)

EMBL; LI2743; AAA31099.1; --
EMBL; LI2743; AAA31098.1; --

This SWISS-PROT entry is copyright. It is produced through a collaboration

CC INHIBITORS IN HUMAN PLASMA.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C25.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; D26470; BAA05484.1; --
DR MEROPS; C25.001; --
DR InterPro; IPR001769; Peptidase_C25.
DR Pfam; PF01364; Peptidase_C25; 1
KW Virulence; Hydrolase; Thiol protease; Calcium; Signal; Zymogen.
FT SIGNAL 1 24
FT PROPEP 25 227
FT CHAIN 228 991
FT CONFLICT 264 265 RT -> TK (IN REF. 2).
FT SEQUENCE 991 AA; 108782 MW; 03EE3F43CEBE2544 CRC64;
SQ
Query Match 7.5%; Score 74.5; DB 1; Length 991;
Best Local Similarity 26.1%; Pred. No. 17;
Matches 23; Conservative 15; Mismatches 27; Indels 23; Gaps 4;
QY 118 FMGNFMDRVPAK--RALRALNMLE-----CGVSRGFLRSNYEVKG-----H 157
DB 247 YEGDKDFDWNQRLTEVKAEDIASPVTAIAQFOVKOEYKEGNLDLYVLLVGDH 306
QY 158 RDVOSTLSPG---DOLYQVIOQSWHYRE 182
DB 307 KDIPAKITPGIKSDQYGVQVGNVDHNE 334
RESULT 8
TEGU_HSVEB
ID TEGU_HSVEB STANDARD; PRT; 3421 AA.
AC 228955;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 01-APR-1993 (Rel. 25, Last annotation update)
DE LARGE TEGUMENT PROTEIN.
GN 24.
OS Equine herpesvirus type 1 (strain Ab4p) (EHV-1).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Varicellovirus.
OX NCBI_TaxID=31520;
RN [1]
RP MEDLINE-92295566; PubMed-1318606;
RX Telford E.A.R., Watson M.S., McBride K., Davison A.J.;
RA "The DNA sequence of equine herpesvirus-1";
RT Virology 189:304-316(1992).
RL
CC -1- FUNCTION: TEGUMENT PROTEIN.
CC -1- SIMILARITY: BELONGS TO FAMILY THAT GROUPS TOGETHER HSV-1 UL36,
CC EHV-1 24, EBV BPLF1, HVS-1 64, VZV 22, AND HCMV UL48.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M86664; AAB02459.1; --
DR PIR; G36797; W2BEB6;
DR HSSP; P04002; 1WFA.
DR SEQUENCE 3421 AA; 367078 MW; 5075EFE4739BB7AC CRC64;
SQ

Query Match 7.4%; Score 73.5; DB 1; Length 3421;
Best Local Similarity 26.6%; Pred. No. 85;
Matches 25; Conservative 12; Mismatches 46; Indels 11; Gaps 3;
QY 89 DGHVYGRGNWIKDHTGPIWNPMSIGITFMGNFMDRVPAKRALRALNLECGVSRGFL 148
DB 65 DGLVTEGRAWTAVAGSKGEA--PCMYSVIV-----ELPNKITVANSANALCCVFSRLYG 115
QY 149 RSNVYKGRHVDQSTLSPGDQLYQVIOQSWHYRE 182
DB 116 DSGFYMHGPDGFQSTQIPARQFFDGV--WKSRSSE 147
RESULT 9
YGJJ_ECOLI
ID YGJJ_ECOLI STANDARD; PRT; 356 AA.
AC P42591;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE HYPOTHETICAL PROTEIN YGJJ PRECURSOR.
GN YGJJ OR B3079.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE-97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12";
RL Science 277:1453-1474(1997).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U18997; AAB57880.1; --
DR EMBL; AE000390; AAC76114.1; --
DR EcoGene; EG12721; YGJJ.
KW Hypothetical protein; Signal; Complete proteome.
FT SIGNAL 1 21 POTENTIAL.
FT CHAIN 22 356 HYPOTHETICAL PROTEIN YGJJ.
SQ SEQUENCE 356 AA; 40099 MW; 51D6089E92848385 CRC64;
Query Match 7.3%; Score 72.5; DB 1; Length 356;
Best Local Similarity 19.4%; Pred. No. 8.4;
Matches 42; Conservative 25; Mismatches 71; Indels 79; Gaps 10;
QY 27 RALPSCSRLGHPVYVVISHTAGSFCSNPSDCEOQARNVQH----- 69
DB 2 KLITAPCRALLALP-----FCYAFSAAGEARPAEHDDTKTPAITSTSSPSFR 49
QY 70 -YHKNELGWCDAVYNFLIGDGHVYE-----GRCWNTKGDTHTGP 107
DB 50 FYGELGVG---CYMDLEGKHKYSDGTVEGKLEMYKSWFGLIYEGCTVQADHDGN 105
QY 108 IWNP-----MSTGIT-FMGNFM--DRVPAKRALRALNLECGVSRGFLRS-NYEVKGH 157
DB 106 AWVPDHSWGFGEGGINFYGYRTNDGTEIMLSLRQSSLDLQWGGDTPDGLGYVPT 165
QY 158 RDVQSTLS-----PGDQLYQVI-----QSWEHY 180
DB 166 RDIMTALKVQNLGNSFRYSVTATPAGHHDESKAWLHF 202

```

-----This SWISS-PROT entry is copyright. It is produced through a collaboration
-----between the Swiss Institute of Bioinformatics and the EMBL outstation -
-----in the European Bioinformatics Institute. There are no restrictions on its
-----use by non-profit institutions as long as its content is in no way
-----modified and this statement is not removed. Usage by and for commercial
-----entities requires a license agreement (See http://www.isb-sib.ch/announce/
-----or send an email to license@isb-sib.ch).
-----
EMBL: M23021; AAA46567.1; --
PIR: A04218; VCBW.
PIR: D33739; VCBWSC.
PDB: 4SBV; 19-APR-89.
InterPro: IPR000937; Viral_coat.
Pfam: PF00729; Viral_coat; 1.
PRINTS: PR00233; ICOSAHEDRAL.
PROSITE: PS00555; ICOSAH_VIR_COAT_S; 1.
Coat protein; 3D-structure; Acetylation.
PROPEP      1      19
CHAIN        20      279      COAT PROTEIN.
MOD_RES      20      279      ACETYLATION.
DOMAIN       20      83      R DOMAIN, INTERACTION WITH RNA.
DOMAIN       84      269      S DOMAIN, VIRION SHELL.
DOMAIN       270      279      P DOMAIN, PROJECTING.
VARIANT      101      101      V -> D (IN STRAIN COMPEA).
VARIANT      245      245      T -> I (IN STRAIN COMPEA).
STRAND       78      81      STRAND
STRAND       86      99      STRAND
STRAND      105      109      STRAND
HELIX        112      115      HELIX
TURN         117      118      TURN
HELIX        119      123      HELIX

```

1. COUNTRY: BELONGS TO THE NIED/NIEK/NIFE/NIEN FAMILY.

1-1- SIMILARITY: BELONGS TO THE NIFD/NIFK/NIFE/NIFN FAMILY.

THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/or_send_an_email_to_license@isb-sib.ch).

EMBL; AF058785; AAC14342.1; -.
InterPro; IPR000318; Nitrogense_compl.
InterPro; IPR000510; Oxidored_nitrogse_1.
Pfam; PF00148; oxidored_nitro; 1.
PROSITE; PS00090; NITROGENASE_1_2; 1.
PROSITE; PS00699; NITROGENASE_1_1; 1.
Oxidoreductase; Nitrogen fixation; Iron-sulfur.
NON_TER 1
NON_TER 280
SEQUENCE 280 AA; 31772 MW; CDF25376AF1056A1 CRC64;

	Very Match	7.2%	Score 72	DB 1	Length 280	8:
Local Similarity	27.6%		Pred. No. 7.2			
Matches	32		Mismatches	43		
Conservative	17					
Indels	24					
Gaps						
54	CNSPDSCEOARVQHYHKNELGWC-----VAYNFLIGEDGHVYEGRWNIKGDH	104				
11111						
122	CNSPGFAGPSQSG-----GHHKINIAWIDQVGTVEPKTSDYVIN-----YVGF-YNIQDQ	173				
105	TGPI--WNPMSTICI-----TFMGNFMDRVPKAKRAL-RAALNLECGVSRGFLRSNYEVK	155				
227	-----GSDVDFPAMHRAHNVLEPCARSAEYICNBLRVR	227				

LT 12
 _HUMAN
 ASAH_HUMAN STANDARD; PRT; 395 AA.
 Q13510;
 15-JUL-1998 (Rel. 36, Created)
 30-MAY-2000 (Rel. 39, Last sequence update)
 20-AUG-2001 (Rel. 40, Last annotation update)
 ACID CERAMIDASE PRECURSOR (EC 3.5.1.23) (ACYLSPHINGOSINE DEACYLASE)
 (N-ACYLSPHINGOSINE AMIDOHYDROLASE) (AC) (PUTATIVE 32 KDA HEART
 PROTEIN) (PHP32).
 ASAH.
 Homo sapiens (Human).
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 NCBI_TaxId=9606;
 [1] SEQUENCE FROM N.A., PARTIAL SEQUENCE, AND VARIANTS V-72; I-93 & K-222.
 TISSUE=Fibroblast, Pituitary, and Urine;
 MEDLINE=97115857; PubMed=8955159;
 Koch J., Gaertner S., Li C.M., Quintern L.E., Bernardo K., Levran O.,
 Schnabel D., Deanick R.J., Schuchman E.H., Sandhoff K.;
 "Molecular cloning and characterization of a full-length complementary
 DNA encoding human acid ceramidase. Identification Of the first
 molecular lesion causing Farber disease.";
 J. Biol. Chem. 271:33110-33115(1996).
 [2] SEQUENCE FROM N.A.
 Churchill J.R., Wieland S.J., Hoffman S., Gallin E.K., Murphy P.M.;
 "A new gene family predicted by a novel human heart cDNA.";
 Mol. Biol. Cell 6:418-418(1995).
 [3] REVISIONS.
 Wieland S.J., Hoffman S., Churchill J.R., Gallin E.K., Murphy P.M.;
 Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
 [4] SEQUENCE FROM N.A.
 Mukherjee A.B.;
 "Human acid ceramidase gene.";

Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases

[5] CHARACTERIZATION: PubMed-7744740;
MEDLINE-95263420; Bernado K., Hurwitz R.; Zenk T., Desnick R.J., Ferlinz K.,
Bernardo K., Hurwitz R.; Zenk T., Desnick R.J., Ferlinz K.,
Schuchman E.H., Sandhoff K.;
Purification, characterization, and biosynthesis of human acid
ceramidase.";
RT J. Biol. Chem. 270:11098-11102(1995).
CC -|- FUNCTION: HYDROLYZES THE SPHINGOLIPID CERAMIDE INTO SPHINGOSINE
CC AND FREE FATTY ACIDS.
CC -|- CATALYTIC ACTIVITY: N-ACYLSPHINGOSINE + H(2)O = A FATTY ACID +
CC SPHINGOSINE.
CC -|- SUBUNIT: HETERODIMER OF ONE ALPHA AND ONE BETA SUBUNIT.
CC -|- SUBCELLULAR LOCATION: LYOSOMAL.
CC -|- TISSUE SPECIFICITY: BROADLY EXPRESSED WITH HIGHEST EXPRESSION IN
CC HEART.
CC -|- PTM: N-GLYCOSYLATED.
CC -|- DISEASE: DEFECTS IN ASAH ARE THE CAUSE OF FARBER'S DISEASE. THIS
CC SPHINGOLIPID DISEASE IS CHARACTERIZED BY SUBCUTANEOUS LIPID-LOADED
CC NODULES, EXCRUCIATING PAIN IN THE JOINTS AND EXTREMITIES, MARKED
CC ACCUMULATION OF CERAMIDE IN LYOSOMES, AND DEATH BY THREE YEARS OF
CC AGE.
CC -|- SIMILARITY: BELONGS TO THE ACID CERAMIDASE FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/>) or send an email to license@isb-sib.ch.

CC	EMBL; U70063; AAC50907.1; ..	1	POTENTIAL.
DR	EMBL; U47674; AAC73009.1; ..	22	ACID CERAMIDASE ALPHA SUBUNIT.
DR	EMBL; AF220175; AAF91230.1; ..	143	ACID CERAMIDASE BETA SUBUNIT.
DR	EMBL; AF220172; AAF91230.1; JOINED.	173	N-LINKED (GLCNAC. . .) (POTENTI
DR	EMBL; AF220173; AAF91230.1; JOINED.	195	N-LINKED (GLCNAC. . .) (POTENTI
DR	MIM; 228000; ..	259	N-LINKED (GLCNAC. . .) (POTENTI
DR	Hydrolase; Glycoprotein; Lysosome; Signal; Disease mutation;	286	N-LINKED (GLCNAC. . .) (POTENTI
KW	Polymorphism.	342	N-LINKED (GLCNAC. . .) (POTENTI
FT	SIGNAL	348	N-LINKED (GLCNAC. . .) (POTENTI
FT	CHAIN	72	M -> V.
FT	CHAIN	93	/FTId=VAR_008860.
FT	CARBOHYD	222	V -> I.
FT	CARBOHYD	222	/FTId=VAR_008861.
FT	CARBOHYD	222	T -> K (IN FARBEN DISEASE).
FT	CARBOHYD	222	/FTId=VAR_008862.
FT	CARBOHYD	21	VSCAVA -> SAVES (IN REF. 2).
FT	CONFLICT	395	AA; 44649 MW; 789F3B936791C128 CRC64;
FT	SEQUENCE		
SQ			

	Query Match	7.2%;	Score 72;	DB 1;	Length 395;
	Best Local Similarity	18.1%;	Pred. No. 11;		
	Matches	43;	Conservative	24;	Mismatches
				75;	Indels
					Gaps
					10;
QY	4 ACALLALIGIATSCSF-----				I V P R S E W R A 28
	: : :				:
Dd	5 SCVALVLLAAAVSCAVAQAHPWTE				C K R S T Y P G P T Y R G A V P W Y I N L D L P P Y K R M H E 64
	:				:
QY	29 LPSECSRL-----				-G H P Y R Y V V I S H T A G S F C N S P D S C Q Q A R N N Q H Y H K 72
	: : :				: : :
Dd	65 LMLDKAPMKVIIVNSLKNMINTFVP				S G K Y Q M V Y D E K L P G L L G N F P G P F E E M K G I A A V T D 124
	:				:
QY	73 NELGWCDVAYNF-----				-L I G E D --G H V Y E G R-----G W N I K G D H T G P T W --- 109


```

Db 125 IPLGEI-ISFNIFVELTICTISVAEDKKGLHNGRNWDGVLGWNIND-----TWIIT 179
QY 110 ---NPMISGIGITFMGNFMDRVPKRALRAALMLLECGVSRGF-----LRSNYEVKG 156
Db 180 EQLKPLAVNLDFORN-----NKTVFKAASSPAGVGYGMLTGFKPLGSLTLNERESING 231

```

RESULT 13

```

LON2_MYXXA
ID LON2_MYXXA STANDARD; PRT; 826 AA.
AC P36774;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE ATP-DEPENDENT PROTEASE LA 2 (EC 3.4.21.53).
GN LON OR BSGA.
OS Myxococcus xanthus.
OC Bacteria; Proteobacteria; delta subdivision; Myxobacteria;
OC Myxococcales; Cystobacterineae; Myxococcaceae; Myxococcus.
OX NCBI_TaxID=34;

```

[1] SEQUENCE FROM N.A.

```

RP STRAIN-M102;
RX MEDLINE-93322335; PubMed-8331083;
RA Gill R.E., Inouye S., Komano T.;
RT "The lon gene is homologous to the lon gene encoding an
ATP-dependent protease and is essential for the development of
Myxococcus xanthus."
RL J. Bacteriol. 175:4545-4549(1993).

```

[2] SEQUENCE FROM N.A., AND SEQUENCE OF 1-9.

```

RP STRAIN-M102;
RX MEDLINE-93322334; PubMed-8331082;
RA Gill R.E., Karlok M., Benton D.;
RT "Myxococcus xanthus encodes an ATP-dependent protease which is
required for developmental gene transcription and intercellular
signaling."
RL J. Bacteriol. 175:4538-4544(1993).
CC -!- FUNCTION: Degrades short-lived regulatory and abnormal proteins.
CC ESSENTIAL FOR FRUITING BODY FORMATION AND DEVELOPMENT.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of large proteins such as globin,
CASEIN AND DENATURATED SERUM ALBUMIN, IN PRESENCE OF ATP.
CC -!- SUBUNIT: HOMOTETRAMER.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -!- DEVELOPMENTAL STAGE: EXPRESSED DURING BOTH VEGETATIVE GROWTH AND
DEVELOPMENT.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S16; ALSO KNOWN AS THE
LON FAMILY OF ATP-DEPENDENT PROTEASES.

```

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to license@isb-sib.ch).

```

CC EMBL; D13204; BAA02491.1; -
CC EMBL; L19301; AAA72018.1; -
CC PIR; A36895; A36895.
CC MEROPS; S16.003; -
CC InterPro; IPR003593; AAA.
CC InterPro; IPR001939; AAA_subfam.
CC InterPro; IPR003111; LON.
CC InterPro; IPR001984; Lon_endopep.
CC Pfam; PF00004; AAA; 1.
CC Pfam; PF02190; LON; 1.
CC PRINTS; PR00830; ENDOLAPTASE.
CC SMART; SM00382; AAA; 1.
CC SMART; SM00464; LON; 1.
CC PROSITE; PS01046; LON_SER; 1.
CC Hydrolase; Serine protease; ATP-binding; Multigene family.

```

```

FT INIT MET 0 0
FT DOMAIN 261 275 GLU-RICH (ACIDIC).
FT NP_BIND 378 385 ATP (POTENTIAL).
FT ACT_SITE 701 BY SIMILARITY.
SQ SEQUENCE 826 AA; 90307 MW; F6765E0D91C948D8 CRC64;

```

Query Match 7.1%; Score 71; DB 1; Length 826;

Best Local Similarity 27.5%; Pred. No. 30; Matches 22; Conservative 13; Mismatches 29; Indels 16; Gaps 3;

```

QY 100 IKGDHTGPIW---NPMISGIGITFMGNFMDRVPKRALRAALMLLECGVSRGFLRSNYEVKG 156
Db 365 LKNDMRGPILCLVPGVGVKTSLGQSVAKATGRKFVRLSL-----GGVRDEABIRG 415

```

```

QY 157 HRDVQSTLSPGDQIQVQIOS 176
Db 416 HRRTYVGALPG----RFIQS 431

```

RESULT 14

```

DPOL_HPBVY
ID DPOL_HPBVY STANDARD; PRT; 832 AA.
AC P03156;

```

```

DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE P PROTEIN [INCLUDES: DNA-DIRECTED DNA POLYMERASE (EC 2.7.7.7); RNA-
DE DIRECTED DNA POLYMERASE (EC 2.7.7.49); RIBONUCLEASE H (EC 3.1.26.4)]
DE (VERSION 1).
GN P.
OS Hepatitis B virus (subtype ayw).
OC Viruses; Retroviruses; Hepadnaviridae; Orthohepadnavirus.
OX NCBI_TaxID=10418;

```

[1] SEQUENCE FROM N.A.

```

RP MEDLINE-81012091; PubMed-399327;
RA Gallibert F., Mandart E., Fitoussi F., Tiollais P., Charnay P.;
RT "Nucleotide sequence of the hepatitis B virus genome (subtype ayw)
cloned in E. coli."
RL Nature 281:646-650(1979).

```

```

CC -!- CATALYTIC ACTIVITY: N DEOXYNUCLEOSIDE TRIPHOSPHATE -
N PYROPHOSPHATE + DNA(N).
CC -!- CATALYTIC ACTIVITY: ENDONUCLEOLYTIC CLEAVAGE TO 5'-PHOSPHO-
MONOSTER.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announcement/
or send an email to license@isb-sib.ch).

```

```

CC EMBL; V01460; -; NOT_ANNOTATED_CDS.
CC PIR; A00702; JDVIVA.
CC InterPro; IPR001462; DNAPol_viral_C.
CC InterPro; IPR000201; DNAPol_viral_N.
CC InterPro; IPR000477; RVTse.
CC Pfam; PF00336; DNA_pol_viral_C; 1.
CC Pfam; PF00242; DNA_pol_viral_N; 1.
CC DR Pfam; PF00078; rvt; 2.
CC DR ProDom; PD000814; DNAPol_viral_C; 1.
CC DR Transferase; RNA-directed DNA polymerase; DNA-directed DNA polymerase;
KW Hydrolase; Nuclease; Endonuclease; DNA replication; DNA-binding.
SQ SEQUENCE 832 AA; 93676 MW; 7AB3AAE58A57D0D6 CRC64;

```

Query Match 7.1%; Score 71; DB 1; Length 832;

Best Local Similarity 23.2%; Pred. No. 31;

Matches 32; Conservative 18; Mismatches 38; Indels 50; Gaps 7;

QY 47 SHTAGSCFNPSDSCBQARNVQH-----YHKNELGWCDA 81

Wed Dec 19 08:45:34 2001

us-09-462-625-2.rspr

Db 236 KMGKVVASITGDSVTAETSAHKAQNLVQCSKSGYIAEEMEKRHGTPHINNFQMR 295

QY 159 DVOSTLSPGDQLY--QVIQSW 177

Db 296 ETMDSLISISEFFFGLELAERK 316

Search completed: December 17, 2001, 07:50:20
Job time: 103 sec

Db 163 THSA-SFGSPYSWEQ---DLOHGAESHQSSGILSRPPVSSLSQSKHRSRLG----- 213
QY 82 YNFLIGEDGHV---YEGRWNIKGDTGPIWNPMSI---GIIFMGNF-----MDRYP 128
Db 214 ---LQSQOGLARROGGSWSIRAGFHTARRPFGVPEPSGGHTTNFASKASCLHQSPV 270
QY 129 KRALRAALNLECGVSRG 146
Db 271 RKAAYPAVSTFEKHSSG 288

RESULT 15
NIFE_METH STANDARD; PRT: 478 AA.
AC O27607;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE NITROGENASE IRON-MOLYBDENUM COFACTOR BIOSYNTHESIS PROTEIN NIFE.
GN NIFE OR MTH1565.
OS Methanobacterium thermoautotrophicum.
OC Archaea; Euryarchaeota; Methanobacteriales; Methanobacteriaceae;
OC Methanothermobacter.
OX NCBI_TaxID=145262;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=DELTA H;
RA MEDLINE=98037514; PubMed=9371463; Deloughery C., Lee H.-M., Dubois J.,
RA Smith D.R., Doucette-Stamm L.A., Blakely D., Cook R., Gilbert K.,
RA Aldredge T., Bashirzadeh R., Keagle P., Lum W., Pothier B., Qiu D.,
RA Harrison D., Hoang L., Wang Y., Wierzbowski J., Gibson R.,
RA Spadafora R., Vicare R., Safer H., Patwell D., Prabhakar S.,
RA Jiwan N., Caruso A., Bush D., Goyal A., Pietrowski S., Church G.M.,
RA McDougall S., Shimer G., Rice P., Noll J., Reeve J.N.;
RA Daniels C.J., Mao J.-I., Rice P., Noll J., Reeve J.N.;
RT "Complete genome sequence of Methanobacterium thermoautotrophicum
RT deltaH: functional analysis and comparative genomics.";
RL J. Bacteriol. 179:7135-7155(1997).
CC -!- FUNCTION: THIS PROTEIN MAY PLAY A ROLE IN THE BIOSYNTHESIS OF
CC -!- PATHWAY: FE-MO COFACTOR BIOSYNTHESIS.
CC -!- SIMILARITY: BELONGS TO THE NIFD/NIFE/NIF/NIFEN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE000916; AAB86039.1;
DR HSPP; P07328; IN2C.
DR InterPro; IPR000318; Nitrogenase_compl.
DR InterPro; IPR000510; Oxidored_nitrognse_1.
DR Pfam; PF01148; oxidored_nitro; 1.
DR PROSITE; PS00699; NITROGENASE_1_1; 1.
DR PROSITE; PS00090; NITROGENASE_1_2; 1.
KW Nitrogen fixation; Complete proteome.
SQ SEQUENCE 478 AA; 52909 MW; AE0285E766E46BFE CRC64;

Query Match 7.0%; Score 70; DB 1; Length 478;
Best Local Similarity 23.4%; Pred. No. 21;
Matches 33; Conservative 28; Mismatches 52; Indels 28; Gaps 10;

QY 63 QARNVQHYHKNELGW---CDVAYNFLIGE--DGHVYEGRWNIKGDTGPIWNPMSI--- 112
Db 178 QSEGRFSFNKS-LGHLQACDVLIDHVGIGDAEYGE-RSYNLVGEFNVAGDLWEIKRL 235
QY 113 SIGITFGNFM-DRYPAK--RALRAALNLECGVSRGFLRSNYEVK-----GHR 158

OM protein - protein search, using sw model

Run on: December 17, 2001, 07:48:37 ; Search time 23.56 Seconds
(without alignments)
1129.948 Million cell updates/sec

Title: US-09-462-625-2

Perfect score: 996

Sequence: 1 MLFACALLALLGLATSCSFI.....TLSPGDQLYQVIOQWHEHYRE 182

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :			
SPTREMBL_17.*			
1:	sp-archaea.*		
2:	sp-bacteria.*		
3:	sp-fungi.*		
4:	sp-human.*		
5:	sp-invertebrate.*		
6:	sp-mammal.*		
7:	sp-mhc.*		
8:	sp-organelle.*		
9:	sp-phage.*		
10:	sp-plant.*		
11:	sp-rodent.*		
12:	sp-virus.*		
13:	sp-vertebrate.*		
14:	sp_unclassified.*		

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES			
Result No.	Score	Query Length	Description
1	996	100.0	182 11 088593 mus musculus
2	838.5	84.2	183 11 09JULN4
3	713	71.6	193 6 09GK12
4	692	69.5	196 4 075594
5	384	38.6	184 5 09V4X2
6	381	38.3	185 5 09V3B7
7	378	38.0	368 4 09HD75
8	373	37.4	190 5 09VV97
9	364	36.5	203 5 09VYX7
10	362.5	36.4	182 5 076537
11	335.5	33.7	208 5 09BLI2
12	334	33.5	345 5 09VXN9
13	324	32.5	337 5 09VSW0
14	322	32.3	186 5 09VS97
15	321.5	32.3	182 5 09V96
16	321.5	32.3	196 5 09XTN0
17	319	31.9	215 5 09VGN3
18	316	31.7	195 5 097369
19	298	29.9	194 5 09BLL1

RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojohori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaudo I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohetsuki S.,
RA Hayashizaki Y.,
RA *Functional annotation of a full-length mouse cDNA collection.*;
RT Nature 409:685-690(2001).
RN [5]
RN SEQUENCE FROM N.A.
RP Strausberg R.;
RA Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
RL -1- FUNCTION: BINDS SPECIFICALLY TO PEPTIDOGLYCAN AND IS INVOLVED IN
CC INNATE IMMUNITY.
CC -1- FUNCTION: THE SOLUBLE FORM TRIGGERS APOPTOSIS IN VITRO.
CC -1- SUBCELLULAR LOCATION: EXISTS IN BOTH SOLUBLE AND MEMBRANE-
CC ASSOCIATED FORMS.
CC -1- TISSUE SPECIFICITY: STRONGLY EXPRESSED IN SPLEEN AND LUNG. ALSO
CC DETECTED IN BRAIN AND THYMUS. IN THE LUNG, EXPRESSED IN THE
CC INTRAALVEOLAR SPACE, IN THE BRAIN, EXPRESSED IN THE PURKINJE CELLS
CC OF THE CEREBELLUM AND IN CERTAIN LAYERS OF NEURONS IN THE
CC HIPPOCAMPUS. ALSO DETECTED IN CELLS FILLING THE SPACE WITHIN THE
CC INTERSTICIAL VILLOS.
CC -1- SIMILARITY: BELONGS TO THE PEPTIDOGLYCAN RECOGNITION PROTEIN
CC FAMILY.
DR EMBL; AF076482; AAC31821.1; -;
DR EMBL; AF193843; AAF06335.1; -;
DR EMBL; AK008335; BAB25611.1; -;
DR EMBL; BC005582; AAH05582.1; -;
DR EMBL; X86374; CAA60133.1; ALT_SEQ.
DR EMBL; Y12088; CAA72803.1; -;
DR MGD; MG1:1345092; Pglyrp.
DR IMMUNE response; Cytokine; Apoptosis; Signal.
KW SIGNAL
FT CHAIN 1 18
FT CHAIN 19 182
FT DISULFID 54 60
FT POTENTIAL.
FT POTENTIAL.
SQ SEQUENCE 182 AA; 20489 MW; 9844E2137F047F14 CRC64;

Query Match 100.0%; Score 996; DB 11; Length 182;
Best Local Similarity 100.0%; Pred. No. 1.2e-95;
Matches 182; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLFACALLALLGLATSCSFIVPRSEWRALPSECSRLGHPVRYVVISHTAGSFCNSPDS 60
Db 1 MLFACALLALLGLATSCSFIVPRSEWRALPSECSRLGHPVRYVVISHTAGSFCNSPDS 60

QY 61 EQQARNVQHYHKNLGCWCDVAYNFLIGEDGHVYEGRGWNIKGDTGPIWNPMSIGITFM 120
Db 61 EQQARNVQHYHKNLGCWCDVAYNFLIGEDGHVYEGRGWNIKGDTGPIWNPMSIGITFM 120

QY 121 NFMDRVPAPKRALRAALNLLCGVSRGFLRSNVEYKGRDVGQSTLSPGDQLYQVIOQSWHY 180
Db 121 NFMDRVPAPKRALRAALNLLCGVSRGFLRSNVEYKGRDVGQSTLSPGDQLYQVIOQSWHY 180

QY 181 RE 182
Db 181 RE 182

RESULT 2
Q9JLN4 PRELIMINARY; PRT; 183 AA.
ID Q9JLN4

AC Q9JLN4;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
DE PEPTIDOGLYCAN RECOGNITION PROTEIN PGRP.
GN PGRP.
OS Rattus norvegicus (rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OC NCBI_TaxID=10116;
RN [1]
RN SEQUENCE FROM N.A.
RP STRAIN=SPRAGUE-DAWLEY; TISSUE=SPLEEN;
RA Rehman A., Teodecki E.E., Krueger J.M.;
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF154114; AAF73252.1; -;
SQ SEQUENCE 183 AA; 20590 MW; 5B9C1B7AA8A2EC21 CRC64;

Query Match 84.2%; Score 838.5; DB 11; Length 183;
Best Local Similarity 84.7%; Pred. No. 2.8e-79;
Matches 155; Conservative 10; Mismatches 17; Indels 1; Gaps 1;

QY 1 MLFACA-LLALLGLATSCSFIVPRSEWRALPSECSRLGHPVRYVVISHTAGSFCNSPDS 59
Db 1 MLFAWAPPALLGLADSCCFVPRSEWKALPSECSGLKKPKPVRYVVISHTAGSFCNSPDS 60

QY 60 CEQARNVQHYHKNLGCWCDVAYNFLIGEDGHVYEGRGWNIKGDTGPIWNPMSIGITFM 119
Db 61 CEQARNVQHYHKNLGCWCDVAYNFLIGEDGHVYEGRGWNIKGDTGPIWNPMSIGITFM 120

QY 120 GNEMDRVPAPKRALRAALNLLCGVSRGFLRSNVEYKGRDVGQSTLSPGDQLYQVIOQSWHY 179
Db 121 GDSHVPAPKRALRAALNLLCGVSRGFLRSNVEYKGRDVGQSTLSPGDQLYQVIOQSWHY 180

QY 180 YRE 182
Db 181 YRE 183

RESULT 3
Q9GK12 PRELIMINARY; PRT; 193 AA.
ID Q9GK12
AC Q9GK12;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
DE PEPTIDOGLYCAN RECOGNITION PROTEIN PRECURSOR.
OS Camelus dromedarius (Dromedary) (Arabian camel).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Tylopoda; Camelidae; Camelus.
OC NCBI_TaxID=9838;
RN [1]
RN SEQUENCE FROM N.A.
RP TISSUE=LACTATING MAMMARY GLAND;
RA Kappeler S.R., Farah Z., Puhan Z.;
RL "Milk as a Source of Camel (Camelus dromedarius) Peptidoglycan
RT Recognition Protein.";
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ131676; CAC19553.1; -;
KW Signal.
FT SIGNAL 1 21
FT CHAIN 22 193
FT POTENTIAL.
FT PEPTIDOGLYCAN RECOGNITION PROTEIN.
SQ SEQUENCE 193 AA; 21377 MW; B6A1BD818030A7CB CRC64;

Query Match 71.6%; Score 713; DB 6; Length 193;
Best Local Similarity 70.6%; Pred. No. 3.2e-66;
Matches 132; Conservative 15; Mismatches 34; Indels 6; Gaps 1;

QY 1 MLFACALLALLGLAT-----SCSFIVPRSEWRALPSECSRLGHPVRYVVISHTAGSFC 54
Db 6 VLLVWALLALLSLGAAREDPACGSIIVPREWRALACRECLTRPVRYVVISHTAGSHC 65

QY 55 NSPDSCEQARNVQHYHKNELGWCDAVYNFLIGEDGHVYEGRWIKGDHTGPIWNPMSI 114
 DB 66 DTPASCAQAQVQSHVNLGWCVDGYNFLIGEDGLVYEGRWIKGAHAGPTWNPISI 125
 QY 115 GITFMGNFMDRVPKRALRAALNLGECVSRGFLRSNVEVKGHRDQVSTLSPGDQLYQVI 174
 DB 126 GISFMGNFMDRVPKRALRAALNLGECVSRGFLRSNVEVKGHRDQVSTLSPGDQLYQVI 185
 QY 175 QSWHYR 181
 DB 186 QWSHYR 192

RESULT 4
 ID 075594 PRELIMINARY; PRT: 196 AA.
 AC 075594;
 DT 01-NOV-1998 (TREMBLrel. 08, Created)
 DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE PEPTIDOGLYCAN RECOGNITION PROTEIN PRECURSOR.
 GN PGLYRP OR PGRP.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A., AND TISSUE SPECIFICITY.
 RC TISSUE=BONE MARROW;
 RX MEDLINE=98374308; PubMed=9707603;
 RA Kang D., Liu G., Lundstroem A., Gelius E., Steiner H.;
 RT "A peptidoglycan recognition protein in innate immunity conserved from
 insects to humans";
 RL Proc. Natl. Acad. Sci. U.S.A. 95:10078-10082(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Lamerdin J.E., McCreedy P.M., Richardson P., Sakaldasis G.,
 RA Burkhardt-Schultz K., Gordon L., Scott D., Johnson G., Stillwagen S.,
 RA Phan H., Velasco N., Do L., Regala W., Terry A., Danganan L.,
 RA Erler A., Christensen M., Georgescu A., Avila J., Attix C.,
 RA Andreise T., Amico-keller G., Coefield J., Duarte S., Lucas S.,
 RA Bruce R., Thomas P., Quan G., Krommiller B., Arellano A., Sanders C.,
 RA Ow D., Nolan M., Trong S., Kobayashi A., Olsen A.S., Carrano A.V.;
 RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Wan T., Zhang W., Cao X.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: BINDS SPECIFICALLY TO PEPTIDOGLYCAN AND IS INVOLVED IN
 CC INNATE IMMUNITY.
 CC -1- FUNCTION: THE SOLUBLE FORM TRIGGERS APOPTOSIS IN VITRO (BY
 CC SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: EXISTS IN BOTH SOLUBLE AND MEMBRANE-
 CC ASSOCIATED FORMS (BY SIMILARITY).
 CC -1- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN BONE MARROW, WEAK
 CC EXPRESSION FOUND IN KIDNEY, LIVER, SMALL INTESTINE, SPLEEN,
 CC THYMUS, PERIPHERAL LEUKOCYTE LUNG AND FETAL SPLEEN.
 CC -1- SIMILARITY: BELONGS TO THE PEPTIDOGLYCAN RECOGNITION PROTEIN
 CC FAMILY.
 CC EMBL; AF076483; AAC31822.1; -;
 CC EMBL; AC007785; AAD38243.1; -;
 CC EMBL; AF242517; AAF99598.1; -;
 CC HSPF; P00806; ILBA.
 CC MIM; 604963; -;
 DR DR Immune response; Cytokine; Apoptosis; Signal.
 KW SIGNAL 1 21
 FT CHAIN 22 196
 FT DISULFID 67 73
 SQ SEQUENCE 196 AA; 21731 MW; D554C51440DC27DC CRC64;

Query Match

69.5%; Score 692; DB 4; Length 196;

Best Local Similarity 67.28; Pred. No. 5e-64;
 Matches 127; Conservative 21; Mismatches 33; Indels 8; Gaps 1;

QY 1 MLFACALLALGLATS-----CSFTVPRSEWRALPSECRSLGRHVPVYVVSHTAGS 52
 DB 6 MLLAWALPSLLRLGAQETEDPACCSPIVPRNWKALASECAQHLSPLRVVVSHTAGS 65
 QY 53 FCNPSDSCQOQARNVQHYHKNELGWCDAVYNFLIGEDGHVYEGRWIKGDHTGPIWNP 112
 DB 66 SCNTPASCOQOQARNVQHYHKNELGWCDAVYNFLIGEDGHVYEGRWIKGDHTGPIWNP 125
 QY 113 SIGITFMGNFMDRVPKRALRAALNLGECVSRGFLRSNVEVKGHRDQVSTLSPGDQLYQ 172
 DB 126 SIGITFMGNFMDRVPKRALRAALNLGECVSRGFLRSNVEVKGHRDQVSTLSPGDQLYH 185
 QY 173 VIOSWEHYR 181
 DB 186 LIQNWPHYR 194

RESULT 5
 ID 09VAX2 PRELIMINARY; PRT: 184 AA.
 AC 09VAX2;
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
 DE CG14745 PROTEIN.
 GN CG14745.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champagne M., Pfeiffer B.D.,
 Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Aghayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballwey R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Dou P.L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Fodor C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heilmann T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclab J.M., Reese M.G.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,

RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pachleb J.M., M.G.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinelt K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J., Yao Q.A.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Zhao Q.,
RA Ye J., Yen R.F., Zaveri J.S., Zhan M., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
RA "The genome sequence of Drosophila melanogaster";
RL Science 287:2185-2195(2000).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE-20559582; PubMed-11106397;
RT Werner T., Liu G., Kang D., Ekengren S., Steiner H., Hultmark D.,
RT "A family of peptidoglycan recognition proteins in the fruit fly
RT Drosophila melanogaster";
RL Proc. Natl. Acad. Sci. U.S.A. 97:13772-13777(2000).
DR EMBL: AE003836; AAF59052.1; -;
DR EMBL: AE003836; AAF59054.1; -;
DR EMBL: AF207542; AAG23736.1; -;
DR HSSP: P00806; IARO.
DR FlyBase: FBgn0033325; CG14746.
DR FlyBase: FBgn0033327; PGRP-Sc1b.
DR SEQUENCE 185 AA; 20395 MW; F23F8D80A33541AC CRC64;
SQ

Query Match 38.3%; Score 381; DB 5; Length 185;
Best Local Similarity 41.7%; Pred. No. 9,1e-32;
Matches 75; Conservative 30; Mismatches 67; Indels 8; Gaps 2;

QY 8 LALLGLATSCS-----FIVPRSEWRALPSCSSRLGHVPVYVISHTAGSCNPDSC 60
DB 5 VALLAVLVCQYMAQGVVYVSKAEGWGKAKWTGLNYLVAIIHHTAGSYCETRAQC 64
QY 61 EQQARNVQVHKHNLGWCVDVAYNFLIGEDGHVYEGRGWIKGDDHTGPIWNPMSITGTFMG 120
DB 65 NAVLSQVQVYHMDSLGWPDIYGNFLIGDGNVYEGRGWNNNGAAAE-WNPYSIGISFLG 123
QY 121 NFMDRVPKAKRALRALNLECGVSGRFLRSNYEYKGRDVGSTLSPGDQLYQVTSWEHY 180
DB 124 NYNWDLEPNMISRAAQLLNDVAVNRGLSSGYILYGHQVSAATECPGTHIWNIRGMSHW 183

RESULT 7
Q9HD75 PRELIMINARY; PRT; 368 AA.
ID Q9HD75
AC Q9HD75;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DE 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE HYPOTHETICAL 40.0 KDA PROTEIN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Zhang W., Wan T., Cao X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF242518; AAF99599.1; -;
DR InterPro: IPR002086; Aldehyde_dehydr.
DR PROSITE: PS00070; ALDEHYDE_DEHYDR_CYS; UNKNOWN_1.
DR Hypothetical protein.
KW SEQUENCE 368 AA; 40020 MW; 1E74970732A5DAFD CRC64;
SQ

Query Match 38.0%; Score 378; DB 4; Length 368;
Best Local Similarity 40.5%; Pred. No. 4,4e-31;
Matches 75; Conservative 33; Mismatches 61; Indels 16; Gaps 4;

RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RL "The genome sequence of Drosophila melanogaster";
RL Science 287:2185-2195(2000).
DR EMBL: AE003836; AAF59051.1; -;
DR HSSP: P00806; IARO.
DR FlyBase: FBgn0033328; CG14745.
DR SEQUENCE 184 AA; 19829 MW; 0F99D04914B07238 CRC64;
SQ

Query Match 38.6%; Score 384; DB 5; Length 184;
Best Local Similarity 42.5%; Pred. No. 4,4e-32;
Matches 77; Conservative 33; Mismatches 63; Indels 8; Gaps 4;

QY 1 MLPACALLLGLATSCSITVPRSEWRALPSCSSRLGHVPVYVISHTAGSCNPDSC 60
DB 11 VLF-CA-QAVLGVY-----IISKSEWGRSATSKTSLANLYSVAVIHHTAGNYCSTKAAC 63
QY 61 EQQARNVQVHKHNLGWCVDVAYNFLIGEDGHVYEGRGWIKGDDHTGPIWNPMSITGTFMG 120
DB 64 ITQLQNTQVYHMDSLGWDITGYNFLIGDGNVYEGRGWNNVGAH-ATNNNSKISIGISFLG 122
QY 121 NFMDRVPKAKRALRALNLECGVSGRFLRSNYEYKGRDVGSTLSPGDQLYQVTSWEHY 180
DB 123 NYNWDLEPNMISRAAQLLNDVAVNRGLSSGYILYGHQVSAATECPGTHIWNIRGMSHW 182
QY 181 R 181
DB 183 K 183

RESULT 6
Q9V3B7 PRELIMINARY; PRT; 185 AA.
ID Q9V3B7
AC Q9V3B7;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DE 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE CG8577 PROTEIN AND CG14746 PROTEIN (PEPTIDOGLYCAN-RECOGNITION PROTEIN-
DE SC1b).
DN GPRP-Sc1b OR (CG8577.3 OR CG8577).
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BERKELEY;
RX MEDLINE-20196006; PubMed-10731132;
RA Adams M.D., Ceiniker S.E., Holt R.A., Evans C.A., Gocayne J.D.F.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Andrews-Pfannkoch C., Baldwin D.,
RA Abril J.F., Agbayani A., An H.-J., Bayraktaroglu L., Beasley E.M.,
RA Ballieu R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Brokstein P., Brotter P.,
RA Borkova D., Botchan M.R., Bouck J., Cadieu E., Center A., Chandra I.,
RA Burtis K.C., Busam D.A., Butler H., Cadiou E., Davies P.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Dou L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fostler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,

```
QY 10 LLGLATSCSFIVPRSEWR-----AL-----PSECSRLGHPVRYVVISHTAGSFCNS 56
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
186 LLGKENC--LAPROKTSKLKAPALSHGLGCEGPRPLSRMTLPFAKYIITHTAGTCNI 243
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
QY 57 PDSCEQOARVQHYHKNELGWCVDVAYNFIAGEDGHVYEGRWNIKGDTGPIWNPMSIGI 116
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 244 SDECRLLVRDIQSYIDRLKSCDGYNFLVQDGAIEYGVGNVQGSST-PGYDDIALGI 302
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
QY 117 TFMGNFMDRVPKRALRALNLLCGVSRGFLRSNRYEVKGRDVOSTLSPGDQLYQVOTQS 176
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 303 TFMGTFGTGPPNAAALEAQAOLIOCAVMKGYLTPTNYLLVGHSDVARTLSPQOALYNIIST 362
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
QY 177 WEHYR 181
| : :
Db 363 WPFEX 367

RESULT 8
Q9VV97 PRELIMINARY; PRT; 190 AA.
ID Q9VV97
AC Q9VV97
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DE 01-MAY-2000 (TREMBlrel. 13, Last annotation update)
DE CG9681 PROTEIN.
GN CG9681.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champagne M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.B.G.,
RA Abrial J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
RA Ballieu R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.C., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphree D.R., Pacleb J.M.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RA *The genome sequence of Drosophila melanogaster.*;
```

RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of *Drosophila melanogaster*";
 RL Science 287:2185-2195(2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-DP CL CN BW;
 RX MEDLINE=2055882; PubMed=11106397;
 RA Werner T., Liu G., Kang D., Ekengren S., Steiner H., Hultmark D.;
 RT "A family of peptidoglycan recognition proteins in the fruit fly
Drosophila melanogaster";
 RL Proc. Natl. Acad. Sci. U.S.A. 97:13772-13777(2000).
 DR EMBL; AF003486; AAF48056.1; -;
 DR EMBL; AF207541; AAG23735.1; -;
 DR EMBL; AF207540; AAG23734.1; -;
 DR FlyBase; FBgn0030310; PGRP-SA.
 SQ SEQUENCE 203 AA; 22260 MW; D200A6EA79C66731 CRC64;

Query Match 36.5%; Score 364; DB 5; Length 203;
 Best Local Similarity 42.5%; Pred. No. 6e-30;
 Matches 79; Conservative 27; Mismatches 62; Indels 18; Gaps 4;

QY 7 LIALGLGTS-----CSFVPRSEWRALPSCSSRLG-----HPVRYVVISHTAGSFC 54
 Db 20 LLAFVSAGKSQRSPANCPTIKLRQWGGKPS-----LGLHYQVRPIRYVVIHHTVTGEC 74
 QY 55 NSPDSCEQOARNVQHYHKNELGWCDAVNFLLIGEDGHVYEGRWGNKIGDHTGPIWNPMSI 114
 Db 75 SGLLKCAEILQNNQAYHONELDFNDISYFNELGNDGIVYEGTGWGLRGAHYG-YNAIGT 133
 QY 115 GIFMGNFMDRVPKRALRAALNLECGVSRGFLRSYEVKGRHVDQSTLSPGDLYQVY 174
 Db 134 GIAFIGNFVDKPLSDAALQAQKLLAGVGQVQGLSESDYALIASQVISTQSPGLTYNEI 193
 QY 175 QSWEHY 180
 Db 194 QEWPHW 199

RESULT 10
 ID 076537 PRELIMINARY; PRT; 182 AA.
 AC 076537;
 DT 01-NOV-1998 (Tremblrel. 08, Created)
 DT 01-NOV-1998 (Tremblrel. 08, Last sequence update)
 DE PEPTIDOGLYCAN RECOGNITION PROTEIN PRECURSOR.
 GN PGRP.
 OS Trichoplusia ni (Cabbage looper).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;
 OC Noctuidae; Noctuidae; Plusiinae; Trichoplusia.
 OX NCBI_TaxID=7111;
 RN [1]
 RP SEQUENCE FROM N.A.; SEQUENCE OF 17-21, AND CHARACTERIZATION.
 RC TISSUE=LARVA;
 RX MEDLINE=98374308; PubMed=9707603;
 RA Kang D., Liu G., Lundstrom A., Gellus E., Steiner H.;
 RT "A peptidoglycan recognition protein in innate immunity conserved from
 insects to humans";
 RL Proc. Natl. Acad. Sci. U.S.A. 95:10078-10082(1998).
 CC -1- FUNCTION: BINDS SPECIFICALLY TO PEPTIDOGLYCAN AND TRIGGERS THE
 CC PROPENOLOXIDASE CASCADE WHICH IS AN IMPORTANT INSECT IMMUNE
 CC DEFENSE MECHANISM.
 CC -1- SUBUNIT: MONOMER (PROBABLE).
 CC -1- TISSUE SPECIFICITY: STRONGLY EXPRESSED IN FAT BODY WITH WEAK
 CC EXPRESSION OBSERVED IN HEMOCYTE. NO EXPRESSION DETECTED IN GUT.
 CC -1- INDUCTION: BY BACTERIAL CHALLENGE.
 CC -1- SIMILARITY: BELONGS TO THE PEPTIDOGLYCAN RECOGNITION PROTEIN
 CC FAMILY.

DR EMBL; AF076481; AAC31820.1; -;
 KW Immune response; Signal.
 FT SIGNAL 1 16
 FT CHAIN 17 182 PEPTIDOGLYCAN RECOGNITION PROTEIN.
 FT DISULFID 18 140 POTENTIAL.
 FT DISULFID 54 60 POTENTIAL.
 SQ SEQUENCE 182 AA; 20572 MW; 56631E762AE34794 CRC64;

Query Match 36.4%; Score 362.5; DB 5; Length 182;
 Best Local Similarity 44.1%; Pred. No. 7.5e-30;
 Matches 71; Conservative 26; Mismatches 63; Indels 1; Gaps 1;

QY 20 IVPSEWRALPSCSSRLGHPVRYVVISHTAGSFCNSPDSCEQOARNVQHYHKNELGWC 79
 Db 20 VYTKDEWDGLTPIHVEYLARPVELVLIQHTVSTCTDACAQIVRNQIYHMDNLNWD 79
 QY 80 VAYNFLLIGEDGHVYEGRWGNKIGDHTGPIWNPMSIGITFMGNFMDRVPKRALRAALNLL 139
 Db 80 IGSFLLIGNGKVEGAGLHVGAHYG-YNRKSIGITFIGNYNDKPTQKSLDALRAL 138

QY 140 ECGVSRGFLRSYEVKGRHVDQSTLSPGDLYQVYQSWEHY 180
 Db 139 RCGVERGHLTANYHIVGHRLISTESPGRLYNEIRWDHF 179

RESULT 11
 ID 09BLL2 PRELIMINARY; PRT; 208 AA.
 AC 09BLL2;
 DT 01-JUN-2001 (Tremblrel. 17, Created)
 DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
 DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
 DE BACTERIOPHAGE T7 LYSOZYME-LIKE PROTEIN 1 (BTL-LP1).
 GN BTL-LP1.
 OS Bombyx mori (Silk moth).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;
 OC Bombycoidea; Bombycidae; Bombyx.
 OX NCBI_TaxID=7091;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Yang J., Tanai K., Yamakawa M.;
 RT "Bacteriophage T7 lysozyme-like protein 1 (BTL-LP1) cDNA";
 RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AB017519; BAB33294.1; -;
 SQ SEQUENCE 208 AA; 23117 MW; DF02FB621AC9AD16 CRC64;

Query Match 33.7%; Score 335.5; DB 5; Length 208;
 Best Local Similarity 40.7%; Pred. No. 5.6e-27;
 Matches 66; Conservative 27; Mismatches 66; Indels 3; Gaps 3;

QY 20 IVPSEWRALPSCSSRLGHPVRYVVISHTA-GSFCNSPDSCEQOARNVQHYHKNELGWC 78
 Db 33 VCSRDGCAVPSKDTPLNKPVPVVIHHTAIPVTCNTTTCMRDMRSMQKYH-NSLGMG 91
 QY 79 DVAYNFLLIGEDGHVYEGRWGNKIGDHTGPIWNPMSIGITFMGNFMDRVPKRALRAALNLL 138
 Db 92 DIGYHFCVGGGVAYEGRWGNVIGIHAGPA-NKLSIGICLIGDMRVETPPAEQLATTKL 150

QY 139 LECGVSRGFLRSYEVKGRHVDQSTLSPGDLYQVYQSWEHY 180
 Db 151 LSTGVEMGAISSDYKLIGHNQAMTTECFGALLEEISTWDNY 192

RESULT 12
 ID 09VXN9 PRELIMINARY; PRT; 345 AA.
 AC 09VXN9;
 DT 01-MAY-2000 (Tremblrel. 13, Created)
 DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
 DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)

CG8995 PROTEIN (PEPTIDOGLYCAN-RECOGNITION PROTEIN-LE).

DE PGRP-LE OR CG8995.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Gelniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champagne M., Pfeiffer B.D.,
RA Wan K.H., Doyle E.G., Helt J., Andrews-Frankoch C., Baldwin D.,
RA Ballwe R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Dodson K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasto P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merklov G., Milshina N.V., Mobarly C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stappleton M., Strong R., Sun E.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Wu D., Yang S., Yao Q.A.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Zhang G., Zhao Q., Zheng L.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RA "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
RP SEQUENCE 345 AA; 39426 MW; 66CC484B54705AD7 CRC64;

Query Match 33.5%; Score 334; DB 5; Length 345;
Best Local Similarity 43.9%; Pred. No. 1.5e-26;
Matches 75; Conservative 27; Mismatches 53; Indels 16; Gaps 5;

QY 18 SFIVPRSEWRA-LPSECSRRGLHPVRYVISHTAGSCFNSDSCQQA-----RVQHY 70
DB 175 SAIIPRSSWLAQKPMDEPLQLPKYKVVILHTA-----TESSEKRAINVLRLHDMQCF 228
QY 71 HKNELGWCDAVNFLLIGDGHVYEGRGNIKGDTHT-GPTWNPMSIGITFMGNFMDRVPAK 129
DB 229 HIESRGWDIAVNFLLIGDGHVYEGRGNIKGDTHT-GPTWNPMSIGITFMGNFMDRVPAK 286

QY 130 RALRAALNLCEGVRGFLRNYEVKGRDVRQSTLSPGDLXQVIOQWEHY 180
DB 287 DALNMCRLNLLARGVEDGHISTDYRLICHQCNSTESPGRRLYEEIQTWPHF 337
RESULT 13
Q9VSWO PRELIMINARY; PRT; 337 AA.
AC Q9VSWO;
DT 01-MAY-2000 (TRENBLrel. 13, Created)
DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)
DE 01-JUN-2000 (TRENBLrel. 14, Last annotation update)
DE CG4437 PROTEIN.
GN CG4437.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Gelniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champagne M., Pfeiffer B.D.,
RA Wan K.H., Doyle E.G., Helt J., Andrews-Frankoch C., Baldwin D.,
RA April J.F., Agbayani A., An H.-J., Bayraktaroglu L., Beasley E.M.,
RA Ballwe R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Dodson K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Durbin K.J., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasto P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merklov G., Milshina N.V., Mobarly C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stappleton M., Strong R., Sun E.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Wu D., Yang S., Yao Q.A.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Zhang G., Zhao Q., Zheng L.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RA "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
RP SEQUENCE 337 AA; 38180 MW; COAC0BC955226C2E CRC64;

Query Match 32.5%; Score 324; DB 5; Length 337;
Best Local Similarity 36.1%; Pred. No. 1.6e-25;
Matches 60; Conservative 34; Mismatches 66; Indels 6; Gaps 3;

QY 20 IVPRESEWRA-LPSECSRRGLHPVRYVISHTAGSCFNSDSCQQAQNVQHYHKNELG 76
DB 229 HIESRGWDIAVNFLLIGDGHVYEGRGNIKGDTHT-GPTWNPMSIGITFMGNFMDRVPAK 286

```

27 ILDRSEWLGEPP--SGKYPHLKLPVSNIIHHHTATGCGCEDVCIYRMKTIQAFHMKSGF 84
QY 77 WCDVAYNFIIGEDGHVYEGRWNGIKDHTGPIWNPMSIGITFMGNFMDRVPKAKRALRAAL 136
Db 85 WVDIGNFIYVGGDQGVYGRGWHQOQVNG-YCAISVSFAFAGTFTFVMEPPARQIEAAK 143
QY 137 NLECGVSRGFLRSNYEVKGRHVDQSTLSPGDQLYQVIOQSWEHYRE 182
Db 144 RLMDGVRHLRLQPDYHIYAHRLQSLPTESPCQKLFELMQNWRFTQ 189

RESULT 14
Q9VS97 PRELIMINARY; PRT; 186 AA.
AC Q9VS97;
DT 01-MAY-2000 (TEMBLrel. 13, Created)
DT 01-MAY-2000 (TEMBLrel. 13, Last sequence update)
DE CG7496 PROTEIN.
GN CG7496.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Vande M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA April J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
RA Balow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Brokstein P., Brottier P.,
RA Borkova D., Botchan M.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Burtis K.C., Busam D.A., Butler H., Davenport L.B., Davies P.,
RA Cherry J.M., Cavley S., Dahlke C., Davenport L.B., Dew I., Dietz S.M.,
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Mays A.D., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Folsler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J., Yao Q.A.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
DR EMBL; AE003530.1;
DR FlyBase; FBgn0035806; CG7496.
SQ SEQUENCE 186 AA; 20123 MW; 09B0C5A7CF4BBF85 CRC64;

Best Local Similarity 41.5%; Pred. No. 1.2e-25;
Matches 68; Conservative 30; Mismatches 64; Indels 2; Gaps 2;

QY 20 IYPRSEWRAL-PSECSSRLGHPVYVYVISHTAGSFNCSFSDSCQOARNVQHYHKNELGWC 78
Db 22 IVTRAENNAKPPNGAIDSMETPLPRAVIAHTAGGACADDVTCQHMNLFQNFQKOKFS 81
QY 79 DVAYNFIIGEDGHVYEGRWNGIKDHTGPIWNPMSIGITFMGNFMDRVPKAKRALRAAL 138
Db 82 DIGYHYLIGNGKVGYSRQSGAFAGPN-NDGSLGIAFAGFEERAPNKEALDRAKEL 140
QY 139 LECGVSGRGFLRSNYEVKGRHVDQSTLSPGDQLYQVIOQSWEHYRE 182
Db 141 LEQAVKQALVEGYKLLGHQVSAATKSPGEALYALIQQWPNWSE 184

RESULT 15
Q9VV96 PRELIMINARY; PRT; 182 AA.
AC Q9VV96;
DT 01-MAY-2000 (TEMBLrel. 13, Created)
DT 01-MAY-2000 (TEMBLrel. 13, Last sequence update)
DE CG9697 PROTEIN.
GN CG9697.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Vande M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA April J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
RA Balow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Brokstein P., Brottier P.,
RA Borkova D., Botchan M.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Burtis K.C., Busam D.A., Butler H., Davenport L.B., Davies P.,
RA Cherry J.M., Cavley S., Dahlke C., Davenport L.B., Dew I., Dietz S.M.,
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Mays A.D., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Folsler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J., Yao Q.A.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
DR EMBL; AE003526; AAF49421.1;

```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 17, 2001, 07:48:37 ; Search time 23.37 Seconds
(without alignments)
576.865 Million cell updates/sec

Title: US-09-462-625-2

Perfect score: 996

Sequence: 1 MLFACALLALLGLATSCSFI.....TLSPGDLQYQVQSWEHYRE 182

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_1101.*

1: /SID88/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SID88/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SID88/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SID88/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SID88/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SID88/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SID88/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SID88/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SID88/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SID88/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SID88/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SID88/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SID88/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SID88/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SID88/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SID88/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SID88/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SID88/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SID88/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SID88/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SID88/gcgdata/geneseq/geneseq/AA2000.DAT.*
22: /SID88/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	996	100.0	182	AA000770	Mouse tag7 clone p
2	957.5	96.1	181	AAW23723	Murine granulocyte
3	692	69.5	191	AA000771	Human tag7 clone p
4	692	69.5	196	AAAB24022	Human PRO1269 prot
5	692	69.5	196	AAAB25583	Htag7 protein enco
6	692	69.5	196	AAAY96964	Chondrosarcoma pep
7	692	69.5	196	AAAY99400	Human PRO1269 (UNQ
8	692	69.5	196	AAAB66149	Protein of the inv
9	676	67.9	190	AAW23722	Bovine granulocyte
10	406	40.8	369	AAAE00693	Human full length
11	406	40.8	375	AAAE00692	Human full length

12	378	38.0	368	21	AAAY96963	Wound healing tiss
13	325.5	32.7	196	19	AAW37837	Amino acid sequenc
14	325.5	32.7	196	19	AAW37835	Amino acid sequenc
15	313.5	31.5	173	19	AAW37834	Recombinant peptid
16	313.5	31.5	173	19	AAW37836	Amino acid sequenc
17	302.5	30.4	363	21	AAAY94863	Human protein clon
18	302	30.3	243	21	AAAY96962	Keratinocyte pepti
19	302	30.3	244	21	AAW76124	Human secreted pro
20	298.5	30.0	530	22	AAV72664	Murine peptidoglyc
21	289	29.0	116	21	AAV64935	Human 5' EST relat
22	181	18.2	683	22	AAAG2931	C glutamic prote
23	180.5	18.1	132	22	AAAE00694	Human Zypal domain
24	154	15.5	174	22	AAV72663	Human peptidoglyc
25	121	12.1	22	20	AAV00773	Mouse tag7 clone p
26	114	11.4	114	22	AAAG1327	Human AFP protein
27	114	11.4	114	22	AAAE00691	Human truncated gr
28	111.5	11.2	53	22	AAW36298	Peptide #10335 enc
29	101	10.1	19	20	AAV00772	Mouse tag7 clone p
30	86	8.6	531	20	AAV05376	Human HCMV inducib
31	86	8.6	537	22	AAAB60496	Human cell cycle a
32	75	7.5	18	20	AAV00775	Human tag7 clone p
33	74.5	7.5	435	22	AAAB49555	Gingipain R. Porp
34	74.5	7.5	493	20	AAW83125	PrtIIR45 Arginine
35	74.5	7.5	737	16	AAW70186	Arg-gingipain-1
36	74.5	7.5	737	18	AAW34846	Arg-gingipain high
37	74.5	7.5	737	21	AAV67395	Arg-gingipain-2 am
38	74.5	7.5	991	16	AAW77313	P. gingivalis haem
39	74.5	7.5	1687	17	AAW96033	Haemagglutinin pro
40	74.5	7.5	1687	19	AAW69495	Arg-gingipain-2 pr
41	74.5	7.5	1704	16	AAW70188	Arg-gingipain high
42	74.5	7.5	1704	18	AAW34843	Arg-gingipain-2 am
43	74.5	7.5	1704	21	AAV67396	PrtR antigenic pro
44	74.5	7.5	1706	18	AAW24786	PrtR antigenic pro
45	73.5	7.4	341	22	AAW93978	Human protein sequ

ALIGNMENTS

RESULT 1

AA000770
ID AA000770 standard; Protein; 182 AA.

AC AA000770;

DT 18-MAY-1999 (first entry)

DE Mouse tag7 clone protein sequence.

Tag7; tumour growth inhibitor; mammalian tumour; carcinoma; sarcoma; melanoma; leukaemia; apoptosis inducer; mouse.

OS Mus sp.

PN WO9902686-A1.

PD 21-JAN-1999.

PF 10-JUL-1998; 98WO-EP04287.

PR 11-JUL-1997; 97US-0893764.

(BOEH) BOEHRINGER INGELHEIM INT GMBH.

Georgiev G, Kiselev S, Ostermann E, Prokhorchouk E;

WPI; 1999-120887/10.

N-PSDB; AAX21819.

New nucleic acid encoding tag7 - used to inhibit tumour growth and induce apoptosis, for treatment of carcinoma, sarcoma, melanoma and leukaemia

PS Claim 19; Fig 1; 138pp; English.

XX This sequence is the murine tag7 of the invention. Cells containing

CC the tag7 DNA sequence are used to express recombinant tag7. Tag7 is used

CC to produce and purify antibodies; to inhibit growth of mammalian tumours,

CC especially for treating carcinoma (of liver, ovary, breast, cervix, lung,

CC prostate, colon/rectum, bladder, testis, stomach, pancreas, mouth, head

CC and neck, squamous cell carcinoma or teratocarcinoma), sarcoma (Raposi's,

CC osteo- or fibro-sarcomas), melanoma or leukaemia; and as a molecular

CC weight marker. The tag7 polypeptide inhibits tumour growth and induces

CC apoptosis. The tag7 coding sequences are also useful as probes for gene

CC mapping and detection of tag7 gene expression, and as primers. Antibodies

CC against tag7 are used as reagents for detecting tag7; as an antagonist of

CC tag7; for isolating tag7 and therapeutically to inhibit or delay tumour

CC metastasis.

XX Sequence 182 AA;

XX

Query Match 100.0%; Score 996; DB 20; Length 182;

Best Local Similarity 100.0%; Pred. No. 1.1e-99;

Matches 182; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLFACALLALGLATSCSFIVPRSEWRALPSECSRLGHPVRYVVISHTAGSFNCSPDSC 60

DB 1 mlfacallallglatcsfivprsewralspsecsrlghpvrivvishtagsfncspdsc 60

QY 61 EQQARNVQHYHKNELGWCVDVAYNFLIGEDGHVYEGRGWNKIGDHTGPIWNPMSIGITFMG 120

DB 61 eqqarnvqhynknelgwcvdvaynfligedghvyegrgwnikgdtgtpiwnpmsigittfmg 120

QY 121 NFMDRVPAKRALRALNLLEGVSRGFLRSNYEVKGRDVGSTLSPGDLYQVIGSWEHY 180

DB 121 nfmdrvpaakraalraalnlllegsvrgflrsnyevkghrdvgstlspgdlyqvigswehy 180

QY 181 RE 182

DB 181 re 182

RESULT 2

AAW23723

ID AAW23723 standard; Protein; 181 AA.

XX

AC AAW23723;

XX

DT 18-FEB-1998 (first entry)

XX

DE Murine granulocyte peptide A precursor (antimicrobial MGP-A).

XX

KW Antimicrobial peptide; antibiotic; antibacterial; antifungal;

KW fungicide; antiprotozoa; protozoacide; antiviral; virucide;

KW murine granulocyte peptide A; MGP-A; preservative; sepsis;

XX endotoxaemia; mouse.

OS Mus musculus.

XX

PH Key

FT Peptide

FT 170..181

FT /label= Mat.peptide

FT /note= "MGP-A antimicrobial peptide (Claim 3)"

XX

PN WO9729765-A1.

XX

PD 21-AUG-1997.

XX

PF 13-FEB-1997; 97WO-US02218.

XX

PR 16-FEB-1996; 96US-0011834.

XX

PA (REGC) UNIV CALIFORNIA.

XX

PI Selsted ME;

XX WPI; 1997-424753/39.

DR N-PSDB; AAT78510.

XX

PT Antimicrobial (poly)peptide(s) from bovine and murine granulocytes -

PT useful therapeutically, as preservatives for food, in water

PT treatment and in agriculture

XX

PS Claim 9; Fig 5; 56pp; English.

XX

CC This protein comprises the precursor of a novel, claimed

CC antimicrobial peptide from murine neutrophils, designated murine

CC granulocyte peptide A or MGP-A (see AAW23725). Its amino acid

CC sequence was deduced from a cDNA clone (see AAT78510) obtained from

CC murine bone marrow. MGP-A and the bovine homologue, BGP-A (see

CC AAW23724), exhibit activity against Gram-positive and gram-negative

CC bacteria, fungi and viruses, specifically Staphylococcus aureus, C.

CC Escherichia coli, Candida albicans, Salmonella typhimurium and C.

CC neoformans (claimed). They can be used in human or veterinary

CC medicine (particularly to treat disorders associated with

CC lipopolysaccharides, e.g. sepsis and endotoxaemia) or as

CC preservatives in food products or in water supplies (claimed).

CC They can also be applied to crops to reduce post-harvest spoilage

CC or expressed in transgenic plants to increase their disease

CC resistance. They have low immunogenicity. Carboxamidated analogues

CC of MGP-A and BGP-A may also be used.

XX

SQ Sequence 181 AA;

XX

Query Match 96.1%; Score 957.5; DB 18; Length 181;

Best Local Similarity 97.8%; Pred. No. 1.6e-95;

Matches 178; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

QY 1 MLFACALLALGLATSCSFIVPRSEWRALPSECSRLGHPVRYVVISHTAGSFNCSPDSC 60

DB 1 mlfacallallglatcsfivprsewralspsecsrlghpvrivvishtagsfncspdsc 60

QY 61 EQQARNVQHYHKNELGWCVDVAYNFLIGEDGHVYEGRGWNKIGDHTGPIWNPMSIGITFMG 120

DB 61 eqqarnvqhynknelgwcvdvaynfligedghvyegrgwnikgdtgtpiwnpmsigittfmg 120

QY 121 NFMDRVPAKRALRALNLLEGVSRGFLRSNYEVKGRDVGSTLSPGDLYQVIGSWEHY 180

DB 121 nfmdrv-rkaalraalnlllegsvrgflrsnyevkghrdvgstlspgdlyqvigswehy 179

QY 181 RE 182

DB 180 re 181

RESULT 3

AAW00771

ID AAW00771 standard; Protein; 191 AA.

XX

AC AAW00771;

XX

DT 18-MAY-1999 (first entry)

XX

DE Human tag7 clone protein sequence.

XX

KW Tag7; tumour growth inhibitor; mammalian tumour; carcinoma; sarcoma;

KW melanoma; leukaemia; apoptosis inducer; human.

XX

OS Homo sapiens.

XX

PN WO9902686-A1.

XX

PD 21-JAN-1999.

XX

PF 10-JUL-1998; 98WO-EP04287.

XX

PR 11-JUL-1997; 97US-0893764.

QY	173	VIOSEHYR	181
	:	:	
Db	186	llgnwphyr	194
RESULT	6		
AAAY96964			
ID	AAAY96964	standard; Protein; 196 AA.	
XX	XX		
XX	AAAY96964;		
XX	31-OCT-2000	(first entry)	
DE	Chondrosarcoma	peptidoglycan recognition protein-like protein.	
XX	Peptidoglycan recognition protein-like protein; PGRP; PGRP-C; regulator		
KW	Chondrosarcoma;	immunosuppressive; antibacterial; anti-apoptotic;	
KW	tumour necrosis factor; TNF;	anti-infectious; mononuclear phagocyte;	
KW	inhibitor; protein co-ordinate	data.	
XX			
OS	Homo sapiens.		
XX	XX		
FH	Key	Location/Qualifiers	
FT	Peptide	1..21	
FT	Protein	/label= Signal_peptide	
FT	Protein	22..196	
FT	Domain	/label= Mature_protein	
FT	Domain	34..117	
FT	Region	/label= PGRP-like_domain	
FT	Region	1..6	
FT	Region	/label= Antigenic_region	
FT	Region	20..29	
FT	Region	/label= Antigenic_region	
FT	Region	33..43	
FT	Region	/label= Antigenic_region	
FT	Region	63..79	
FT	Region	/label= Antigenic_region	
FT	Region	99..112	
FT	Region	/label= Antigenic_region	
FT	Region	133..146	
FT	Region	/label= Antigenic_region	
FT	Region	160..165	
FT	Region	/label= Antigenic_region	
FT	Region	168..181	
FT	Region	/label= Antigenic_region	
FT	Region	190..196	
FT	Region	/label= Antigenic_region	
XX	WO200039327-A1.		
PN	06-JUL-2000.	99WO-US30736.	
XX	22-DEC-1999;	98US-0113809.	
XX	23-DEC-1998;	(HUMA-) HUMAN GENOME SCI INC.	
XX	PA	Rosen CA, Ruben SM, Yöung PE, Olsen HS;	
XX	PI	WPI; 2000-452414/39.	
XX	DR	N-PSDB; AAA51719.	
XX	PT	polynucleotide encoding peptidoglycan recognition protein-like proteins (PGRP)	
XX	PT	antibodies specific to it useful for preventing, treating conditions	
XX	PT	e.g. endotoxemic shock and auto-immune disorders and infections in mammal	
XX	PS	Claim 11; Fig 3; 191pp; English.	
XX	CC	Novel human peptidoglycan recognition protein-like proteins (PGRP)	
XX	CC	expressed by keratinocytes (PGRP-K), wound healing tissues (PGRP-W)	
XX	CC	or chondrosarcoma tissue (PGRP-C) are disclosed. The PGRP DNA and	
XX	CC	proteins are useful for preventing, treating or ameliorating a medical	

28-OCT-1998; 98US-0106033.
PR 28-OCT-1998; 98US-0106178.
PR 29-OCT-1998; 98US-0106248.
PR 29-OCT-1998; 98US-0106384.
PR 29-OCT-1998; 98US-0108500.
PR 30-OCT-1998; 98US-0106464.
PR 03-NOV-1998; 98US-0106856.
PR 03-NOV-1998; 98US-0106902.
PR 03-NOV-1998; 98US-0106905.
PR 03-NOV-1998; 98US-0106919.
PR 03-NOV-1998; 98US-0106932.
PR 03-NOV-1998; 98US-0106934.
PR 10-NOV-1998; 98US-0107783.
PR 17-NOV-1998; 98US-0108775.
PR 17-NOV-1998; 98US-0108779.
PR 17-NOV-1998; 98US-0108787.
PR 17-NOV-1998; 98US-0108788.
PR 17-NOV-1998; 98US-0108801.
PR 17-NOV-1998; 98US-0108802.
PR 17-NOV-1998; 98US-0108806.
PR 17-NOV-1998; 98US-0108807.
PR 17-NOV-1998; 98US-0108867.
PR 17-NOV-1998; 98US-0108925.
PR 18-NOV-1998; 98US-0108848.
PR 18-NOV-1998; 98US-0108849.
PR 18-NOV-1998; 98US-0108850.
PR 18-NOV-1998; 98US-0108851.
PR 18-NOV-1998; 98US-0108852.
PR 18-NOV-1998; 98US-0108858.
PR 18-NOV-1998; 98US-0108904.
XX
PA (GETH) GENENTECH INC.
XX Baker K, Goddard A, Gurney AL, Smith V, Watanabe CK, Wood WI;
PI N-PSDB; AAA37082.
DR WPI: 2000-237871/20.
XX
XX New mammalian DNA sequences encoding transmembrane, receptor or
PT secreted PRO polypeptides, useful for screening of potential peptide or
PT small molecule inhibitors of the relevant receptor/ligand interactions
XX
PS Claim 12; Fig 122; 773pp; English.
XX
XX AAA37022 to AAA37144 encode the new isolated human transmembrane,
CC receptor or secreted PRO polypeptides given in AAY99340 to AAY99462. The
CC transmembrane and receptor PRO proteins can be used for screening of
CC potential peptide or small molecule inhibitors of the relevant
CC receptor/ligand interactions. The polypeptides and nucleotide sequences
CC encoding then have various industrial applications, including uses as
CC pharmaceutical and diagnostic agents. AAA37145 to AAA37330 represent
CC PCR primers and hybridisation probes used in the isolation of the PRO
CC polypeptides from the present invention.
XX
SQ Sequence 196 AA;
Query Match 69.5%; Score 692; DB 21; Length 196;
Best Local Similarity 67.2%; Pred. No. 8.8e-67;
Matches 127; Conservative 21; Mismatches 33; Indels 8; Gaps 1;
QY 1 MLFACALLALLGLATS-----CSFIVPRSEWRALPSECSSRLGHPVRYVVISHTAGS 52
DB 6 mllawlpallrlgaagetedpaccspivprnewkalasecaghlslplryvvshtags 65
QY 53 FCNSPDCSQOARNOVQHYKNEIGWCDVAYNFIIGDGHVYEGRWNIKGDHTGPIWNPM 112
DB 66 scntpascqqqarnvqhyhmktlgwcdvgnfligedgivyegrgwnftgahsghlwnpm 125
QY 113 SIGITFMGNFMDVPKAKRALNLLLECGRVGRFLRSNYEVKGRHDVOSTLSPGDQLYQ 172
DB 126 sigisfmgnymdrvptpqalraaagglacgacgalarlnsyvlgkhrdvtqlspgqlyh 185

QY 173 VIQSWHYR 181
DB 186 liqnwphyr 194
RESULT 8
AAB66149
ID AAB66149 standard; protein; 196 AA.
XX
AC AAB66149;
XX
DT 02-APR-2001 (first entry)
DX Protein of the invention #61.
DE Secreted; transmembrane; gene therapy.
XX Unidentified.
OS
XX WO200078961-A1.
XX
XX 28-DEC-2000.
XX 18-FEB-2000; 2000WO-US04342.
XX 23-JUN-1999; 99US-0141037.
XX 20-JUL-1999; 99US-0144758.
XX 26-JUL-1999; 99US-0145698.
XX 01-SEP-1999; 99WO-US20111.
XX 29-OCT-1999; 99US-0162506.
XX 30-NOV-1999; 99WO-US28313.
XX 02-DEC-1999; 99WO-US28551.
XX 16-DEC-1999; 99WO-US30095.
XX 05-JAN-2000; 2000WO-US00219.
XX 06-JAN-2000; 2000WO-US00376.
XX
XX (GETH) GENENTECH INC.
XX Baker KP, Botstein D, Desnoyers L, Baton DL, Ferrara N, Fong S;
PI Gao W, Goddard A, Godowski PJ, Grimaldi CJ, Gurney AL, Hillan KJ;
PI Pan J, Paoni NF, Roy MA, Smith V, Stewart TA, Tumas D;
PI Watanabe CK, Williams PM, Wood WI;
XX
XX WPI: 2001-071395/08.
XX
XX Secreted and transmembrane proteins and nucleic acids designated PRO,
PT useful as hybridization probes, in chromosome and gene mapping, and gene
PT therapy -
XX
PS Claim 1; Fig 122; 787pp; English.
XX
XX The present invention relates to secreted and transmembrane proteins.
CC These proteins and the DNA encoding them may be used as hybridization
CC probes, in chromosome and gene mapping and in the generation of
CC anti-sense RNA and DNA. They may also be used to generate either
CC transgenic animals or knockout animals which are in turn useful for
CC development and screening of therapeutically useful reagents.
CC The nucleic acids may also be used in gene therapy.
XX
SQ Sequence 196 AA;
Query Match 69.5%; Score 692; DB 22; Length 196;
Best Local Similarity 67.2%; Pred. No. 8.8e-67;
Matches 127; Conservative 21; Mismatches 33; Indels 8; Gaps 1;
QY 1 MLFACALLALLGLATS-----CSFIVPRSEWRALPSECSSRLGHPVRYVVISHTAGS 52
DB 6 mllawlpallrlgaagetedpaccspivprnewkalasecaghlslplryvvshtags 65
QY 53 FCNSPDCSQOARNOVQHYKNEIGWCDVAYNFIIGDGHVYEGRWNIKGDHTGPIWNPM 112
DB 66 scntpascqqqarnvqhyhmktlgwcdvgnfligedgivyegrgwnftgahsghlwnpm 125

Qy	113	SIGTFPGNWDVPKARALRAAALLLECGVSRGFLRSNYEYVGHRDVTSLSPGDOLYQ	172
		: : : : : : : :	
Db	126	sigsfmgnyndrvptpqaraacgllacvagaalrsnyvlkghrdvqrtlspgnqlyh	185
Qy	173	VIOSEWYR	181
		: :	
Db	186	liapwphyr	194

RESULT	9	
AAW23722		Antimicrobial peptide; antibiotic; antibacterial; antifungal;
ID	AAW23722	standard; Protein; 190 AA.
XX		
XX	AAW23722;	
XX		
XX		
DT	18-FEB-1998	(first entry)
XX		
DE		Bovine granulocyte peptide A precursor (antimicrobial BGP-A).
XX		
KW		Antimicrobial peptide; antibiotic; antibacterial; antifungal;
KW		fungicide; antiprotozoa; protozoacide; antiviral; virucide;
KW		bovine granulocyte peptide A; BGP-A; preservative; sepsis;
KW		endotoxaemia; cattle.

OS Bos taurus.

PH	Key	Location/Qualifiers
FT	Peptide	1..21
FT		/label= sig_peptide
FT	Peptide	22..177
FT		/label= Propeptide
FT	Peptide	178..190
FT		/label= Mat_peptide
FT		/note= "BCP-A antimicrobial peptide (Claim 2)"

XX
PN WO9729765-A1.

PD 21-AUG-1997.

PF 13-FEB-1997; 97WO-US02218.

PR 16-FEB-1996; 96US-0011834.

PA (REGC) UNIV CALIFORNIA.

PI Selsted ME;

DR WPI; 1997-424753/39.

XX
1900, 1911/1900.

antimicrobial (poly)peptide(s) from bovine and marine granulocytes useful therapeutically, as preservatives for food, in water

XX
 11
 CREEDENCE AND IN APPRECIATION[illegible]

CC This protein comprises the precursor (see AMW3722) of a novel,
CC claimed antimicrobial peptide from bovine neutrophils, designated

CC bovine granulocyte peptide A (see AAWZ3/24): ITS amino acid
CC sequence was deduced from a cDNA clone (see AAT78509) obtained from

CC bovine bone marrow. BGF-A and the murine homologue, MGF-A (see
CC AAW23725), exhibit activity against Gram-positive and Gram-negative

CC bacteria, fungi and viruses, specifically *Staphylococcus aureus*,
CC *Escherichia coli*, *Candida albicans*, *Salmonella typhimurium* and C.

neurofibrins (crabmeat). They can be used in human or veterinary medicine (particularly to treat disorders associated with

CC lipopolysaccharides, e.g.: sepsis and endotoxaemia) or as
CC preservatives in food products or in water supplies (claimed).

CC They can also be applied to crops to reduce post-harvest spoilage
CC or expressed in transgenic plants to increase their disease

resistance. They have low immunogenicity. Carboxamidated analogues of BGP-A and MGP-A may also be used.

XX	SQ
Sequence	190 AA;
Query Match	67.9%; Score 676; DB 18; Length 190;
Best Local Similarity	68.3%; Pred. No. 4.5e-65;
Matches 123;	Conservative 19; Mismatches 36; Indels 2; Gaps
Oy	4 ACALLLGL--ATSCSFIVPRSEWRALPSECSSRLGHPVRYVVISHTAGSFCNSPDSCSCE 61 : Db
Dd	9 awlllallglgaadcgslvsrgkwalaskcsgrlrqrvryvvshstagsvcntpasqc 68 :
Oy	62 QOARNVOHYHKNELGWCDYAYNELIGEDGHYYEGRWNIKGDOHTGPINWPMSTGITFMGN 121 :
Dd	69 rqaqvdyvnyvrgegwcdyvnlgedgivyegrwntlgahsgptwnplagisfmgn 128 :
Oy	122 FMDRVPKRALRAALNLECGVSRGFSLRSNYEKGHRDVQSITLSPGDQLYQVIQTWEHYR 181 :
Dd	129 ymhrrvpasalaarsillacaaargyltpnyevghrdvgqltspdeykliiqwphyr 188 :
RESULT	10
AEE00693	
ID	AEE00693 standard; Protein; 369 AA.
AC	AEE00693;
XX	
DT	02-JUL-2001 (first entry)
DE	Human full length granulocyte peptide homolog Zgal protein #2.
KW	Human; granulocyte peptide A; GP-A; Zgal; cytostatic; antiinflammatory; KW vulnery; dermatological; anti-microbial; gastrointestinal disease; KW pulmonary; dental carries; periodontal disease; gene therapy; AIDS; KW Acquired Immune Deficiency Syndrome; epithelial wound; cystic fibrosis; KW infection; urinary tract; respiratory; vaginal; lung; skin; cancer; KW ovarian; rectal; chromosome 1. XX OS Homo sapiens.
XH	
FH	
Key	Location/Qualifiers
Peptide	1..17
Protein	/label= Signal_peptide
FT	18..369
FT	/label= Mature_full_length_Zgal_protein.#2
FT	/note= "This region functions as an immunogenic epitope"
Region	20..29
Domain	/label= Immunogenic_epitope
FT	52..183
FT	/label= Domain_1
Region	84..90
FT	/label= Immunogenic_epitope
FT	/note= "this region is specifically claimed in claim 18"
Region	92..97
FT	/label= Hydrophillic_region
Region	105..110
FT	/label= Immunogenic_epitope
Region	117..122
FT	/label= Hydrophillic_region
Region	150..155
FT	/label= Hydrophillic_region
Region	150..157
FT	/label= Immunogenic_epitope
Region	151..156
FT	/label= Hydrophillic_region
Region	152..157
FT	/label= Hydrophillic_region
Region	184..209
FT	/label= Linker
Region	190..202
FT	/label= Immunogenic_epitope
Region	205..223
FT	/label= Immunogenic_epitope

FT Domain /note= "This region is specifically claimed in claim 18"
 FT 210..340
 FT /label= Domain_2
 FT Region 217..223
 FT /label= Immunogenic_epitope
 FT Region 239..249
 FT /note= "This region is specifically claimed in claim 18"
 FT /label= Immunogenic_epitope
 FT Region 261..267
 FT /label= Immunogenic_epitope
 FT Region 291..297
 FT /label= Immunogenic_epitope
 FT Region 346..354
 FT /label= Immunogenic_epitope
 XX
 PN WO200129224-A2.
 XX
 XX 26-APR-2001.
 PD
 XX 20-OCT-2000; 2000WO-US29177.
 PF
 XX 20-OCT-1999; 99US-0160712.
 PR 12-JUL-2000; 2000US-0218070.
 XX
 XX (ZYMO) ZYMOGENETICS INC.
 PA Conklin DC, Adler DA, Fox BA;
 XX WPI; 2001-290918/30.
 PI N-PSDB; AAD04006.
 DR
 XX
 XX New granulocyte peptide homolog, zgpal polypeptide, for research
 PT applications, diagnosis and treatment of cancer, periodontal,
 PT gastrointestinal disease, urinary tract, skin and lung infections
 PT
 PT Claim 14; Page 109-110; 114pp; English.
 PS
 XX The present sequence is human full length granulocyte peptide (GP-A)
 CC homolog, Zgpal protein. Zgpal gene is located on human chromosome 1.
 CC Zgpal polypeptides are useful for producing antibodies which are useful
 CC for detecting cancer. Zgpal polypeptides having anti-microbial activity
 CC are useful for treating dental caries, periodontal disease, thrush, are
 CC detecting cancer. Zgpal polypeptides having anti-microbial activity are
 CC useful for treating dental caries, periodontal disease, thrush,
 CC gastrointestinal disease, urinary tract infections, vaginal infections,
 CC respiratory infections, acquired immune deficiency syndrome (AIDS) and
 CC lung infections associated with cystic fibrosis and prevention of
 CC infection in skin and other epithelial wounds. Zgpal-cytokine fusion
 CC proteins are useful for enhancing in vivo killing of target tissues
 CC (epithelial cancers, and more specifically lung, ovarian and rectal
 CC cancers). Zgpal polypeptides, fragments, fusion proteins or agonists are
 CC useful in in vitro studies of exogenous microorganism infections such as
 CC bacterial, viral or fungal infection and also to study epithelial cell
 CC defensin induction in cell culture. Zgpal antibodies, polynucleotides and
 CC polypeptides are useful for detection of zgpal polypeptide, mRNA or
 CC anti-zgpal antibodies, thus serving as markers for detecting genetic
 CC diseases or cancers. Zgpal sequences are useful as diagnostics in
 CC forensic DNA profiling and as probes or primers to clone 5' non-coding
 CC regions of a zgpal gene. Zgpal antibodies are useful for tagging cells
 CC that express zgpal, for screening expression libraries and as
 CC neutralizing antibodies or as antagonists to block zgpal activity
 CC in vitro and in vivo. Zgpal gene is also useful in gene therapy.
 XX
 XX Sequence 369 AA;
 SQ
 Query Match 40.8%; Score 406; DB 22; Length 369;
 Best Local Similarity 43.8%; Pred. No. 1.7e-35;
 Matches 74; Conservative 33; Mismatches 60; Indels 2; Gaps 2;
 QY 13 LATSCSFVPRSEWRALPSCSSRLGHPVRVYVHTAGSFCNSPDSCQOQARNVQHYK 72
 DB | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
 202 lktacpgvprsvwgarethc-prmtipakygihtagrtcnisdecrllvrdiqsfyl 260

QY 73 NELGWCDAVNFLLIGEDGHVYEGRWIKGDHTGPIWNPMISIGITFMGNFMDRVPKRAL 132
 : | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
 Db 261 drlkscdigynflvgqgalyegvwnvgssst-pgyddialgitfmgtftgipppnaaal 319
 QY 133 RAALNLECGVSRGFLSRNVEVKGRDQVSTLSPGDOLYQVIOQSEHYR 181
 | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
 Db 320 eaqdlqcamkvgytlpnyllvghsdvartlspgqalynliistwphfk 368
 RESULT 11
 AAE00692
 ID AAE00692 standard; Protein: 375 AA.
 XX
 XX AAE00692;
 AC
 XX 02-JUL-2001 (first entry)
 DT
 XX Human full length granulocyte peptide homolog Zgpal protein #1.
 DE
 XX Human; granulocyte peptide A; GP-A; Zgpal; cytostatic; antiinflammatory;
 XX vulnary; dermatological; anti-microbial; gastrointestinal disease;
 KW pulmonary; dental caries; periodontal disease; gene therapy; AIDS;
 KW Acquired Immune Deficiency Syndrome; epithelial wound; cystic fibrosis;
 KW infection; urinary tract; respiratory; vaginal; lung; skin; cancer;
 KW ovarian; rectal; chromosome 1.
 XX
 XX Homo sapiens.
 OS
 XX
 XX Key Location/Qualifiers
 FH 1..17
 FT Peptide /label= Signal_peptide
 FT Protein 18..375
 FT /label= Mature_full_length_zgpal_protein_#1
 FT /note= "This region functions as an immunogenic epitope"
 FT Region 47..52
 FT /label= Immunogenic_epitope
 FT /note= "This region is specifically claimed in claim 18"
 FT Domain 58..189
 FT /label= Domain_1
 FT Region 98..103
 FT /label= Hydrophilic_region
 FT Region 123..128
 FT /label= Hydrophilic_region
 FT Region 152..157
 FT /label= Hydrophilic_region
 FT Region 156..161
 FT /label= Hydrophilic_region
 FT Region 157..162
 FT /label= Hydrophilic_region
 FT Region 158..163
 FT /label= Hydrophilic_region
 FT Region 190..215
 FT /label= Linker
 FT Domain 216..346
 FT /label= Domain_2
 FT
 FT WO200129224-A2.
 PN
 XX
 XX 26-APR-2001.
 PD
 XX 20-OCT-2000; 2000WO-US29177.
 PF
 XX 20-OCT-1999; 99US-0160712.
 PR 12-JUL-2000; 2000US-0218070.
 XX
 XX (ZYMO) ZYMOGENETICS INC.
 PA Conklin DC, Adler DA, Fox BA;
 XX WPI; 2001-290918/30.
 PI N-PSDB; AAD04004.
 XX

PT New granulocyte peptide homolog, zgpal polypeptide, for research
PT applications, diagnosis and treatment of cancer, periodontal,
PT gastrointestinal disease, urinary tract, skin and lung infections
XX
PS Claim 14; Page 105-106; 114pp; English.
XX
CC The present sequence is human full length granulocyte peptide (GP-A)
CC homolog, zgpal protein. Zgpal gene is located on human chromosome 1.
CC zgpal polypeptides are useful for producing antibodies which are useful
CC for detecting cancer. Zgpal polypeptides having anti-microbial activity
CC are useful for treating dental carries, periodontal disease, thrush,
CC detecting cancer. Zgpal polypeptides having anti-microbial activity are
CC useful for treating dental carries, periodontal disease, thrush,
CC gastrointestinal disease, urinary tract infections, vaginal infections,
CC respiratory infections, acquired immune deficiency syndrome (AIDS) and
CC lung infections associated with cystic fibrosis and prevention of
CC infection in skin and other epithelial wounds. Zgpal-cytokine fusion
CC proteins are useful for enhancing in vivo killing of target tissues
CC (epithelial cancers, and more specifically lung, ovarian and rectal
CC cancers). Zgpal polypeptides, fragments, fusion proteins or agonists are
CC useful in in vitro studies of exogenous microorganism infections such as
CC bacterial, viral or fungal infection and also to study epithelial cell
CC defensin induction in cell culture. Zgpal antibodies, polynucleotides and
CC polypeptides are useful for detection of zgpal polypeptide, mRNA or
CC anti-zgpal antibodies, thus serving as markers for detecting genetic
CC diseases or cancers. Zgpal sequences are useful as diagnostics in
CC forensic DNA profiling and as probes or primers to clone 5' non-coding
CC regions of a zgpal gene. Zgpal antibodies are useful for tagging cells
CC that express zgpal, for screening expression libraries and as
CC neutralizing antibodies or as antagonists to block zgpal activity
CC in vitro and in vivo. Zgpal gene is also useful in gene therapy.
XX
SQ Sequence 375 AA;

Query Match 40.8%; Score 406; DB 22; Length 375;
Best Local Similarity 43.8%; Pred. No. 1.7e-35;
Matches 74; Conservative 33; Mismatches 60; Indels 2; Gaps 2;
QY 13 LATSCSFTVPSEWALSECSSRLGHVRYVVISHTAGSCFNPSDCEQARNVQHYHK 72
Db 208 lkkacpgvprsvwgarethc-prmtlpakgyilhtagrtcnisdecrllvrdiqsfyi 266
QY 73 NELGCDVAYNFLIGEDGHVYEGRWNTKGDHTGPIWNPMSIGITFMGNFMDRVPKRAL 132
Db 267 drlkscdigynflvgdgaivgynvqgssst-pgyddialgitfmgftgipppnaaal 325
QY 133 RAALNLECGVSRGFLRSNYEVKGRDVOSTLSPGDQLYQVYQSWHYR 181
Db 326 eaaqqliqcamvkyltpnyllvghsdvartlspgalylniilstwphfk 374

RESULT 12
AA996963
ID AA996963 standard; Protein: 368 AA.
XX
AC AA996963;
XX
DT 31-OCT-2000 (first entry)
XX
DE Wound healing tissue peptidoglycan recognition protein-like protein.
KW Peptidoglycan recognition protein-like protein; PGRP; PGRP-W; regulator;
KW wound healing tissue; immunosuppressive; antibacterial; anti-apoptotic;
KW tumour necrosis factor; TNF; anti-infectious; mononuclear phagocyte;
KW inhibitor; protein co-ordinate data.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Peptide 1..17
FT /label= signal_peptide
FT Protein 18..368

FT Region /label= Mature_protein
FT 17..32 /label= Antigenic_region
FT Region 40..58 /label= Antigenic_region
FT Domain 52..135 /label= PGRP-like_domain
FT Region 82..99 /label= Antigenic_region
FT Region 104..111 /label= Antigenic_region
FT Region 116..125 /label= Antigenic_region
FT Region 150..159 /label= Antigenic_region
FT Region 174..182 /label= Antigenic_region
FT Region 186..207 /label= Antigenic_region
FT Region 214..225 /label= Antigenic_region
FT Region 237..252 /label= Antigenic_region
FT Region 259..268 /label= Antigenic_region
FT Region 290..300 /label= Antigenic_region
FT Region 344..355 /label= Antigenic_region
FT Region 364..368 /label= Antigenic_region
XX
PN WO200039327-A1.
XX
PD 06-JUL-2000.
XX
PF 22-DEC-1999; 99WO-US30736.
XX
PR 23-DEC-1998; 98US-0113809.
XX (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Ruben SM, Young PE, Olsen HS;
XX
DR WPI: 2000-452414/39.
DR N-PSDB; AAA51718.
XX
PT Polynucleotide encoding peptidoglycan recognition protein-like protein,
PT antibodies specific to it useful for preventing, treating conditions
XX e.g. endotoxemic shock and auto-immune disorders and infections in mammal
PS Claim 11; Fig 2A-B; 191pp; English.
XX
CC Novel human peptidoglycan recognition protein-like proteins (PGRP)
CC expressed by keratinocytes (PGRP-K), wound healing tissues (PGRP-W)
CC or chondrosarcoma tissue (PGRP-C) are disclosed. The PGRP DNA and
CC proteins are useful for preventing, treating or ameliorating a medical
CC condition in a mammal (claimed). PGRP is useful in augmenting the immune
CC system in such areas as immune recognition, antigen presentation and
CC immune system activation. Antibodies or antagonists directed against
CC these proteins may be useful in reducing or eliminating disorders
CC associated with tumour necrosis factor (TNF) and TNF-like cytokines,
CC such as endotoxemic shock and autoimmune disorders and for treating
CC infectious diseases including silicosis, sarcoidosis and idiopathic
CC pulmonary fibrosis.
XX
SQ Sequence 368 AA;

Query Match 38.0%; Score 378; DB 21; Length 368;
Best Local Similarity 40.5%; Pred. No. 1.8e-32;
Matches 75; Conservative 33; Mismatches 61; Indels 16; Gaps 4;

	Query Match	32.7%	Score 325.5;	DB 19;	Length 196;
	Best Local Similarity	38.8%;	Pred. No. 3.6e-27;		
	Matches	71;	Conservative	29;	Mismatches 80; Indels 3; Gaps
Qy	1	MLFACALLALL-GLATSCSFIVPRSEWRALPSCSSRELGHDPYRVYVISHTAGSCFNCPDS	59		
	:	:	:	:	:
	:	:	:	:	:
Db	8	vvlalalsllteiaadcd-vvskkqgdglphvhsylarpslvigqvtvpcftcdag	66		
	:	:	:	:	:
Qy	60	CEQQAARVQVYHKNELGWCQDVAVNFLIGDGHVYEGRGNNIKGDHPTWNPMSIGITGM	119		
	:	:	:	:	:
Db	67	ceelvrnltqtnhmealgdywdgpfslvgvgngkyvyesgwllhvgahkyg-ynsrzslqvaf	125		
	:	:	:	:	:
Qy	120	GNFMDRVPAKRALRAAALLNLECGYSGRGLFSNRYEVKGRHDVQSTLSFGDQLYQVIOGSWEH	179		

Query Match	32.78;	Score	325.5;	DB	19;	Length	196;
Best Local Similarity	38.8;	Pred. No.	3.6e-27;				
Matches	71;	Conservative	29;	Mismatches	80;	Indels	3;
Gaps							
3:							
Qy	1	MLFACALIALI-LGATSCSFIVPSEWRALPESESSRLGHPVRVYVVISHTAGTSCFCNSPDS	59				
Db	8	vvlalalslliteaadd-vvskkqgdglipvhvsylarpvslvivqhtvtpfertdag	66				
Qy	60	CEQQAQNVQHHKELGWCDVAYFLICEDGHVYEGCGWNKIGDHTGFWNPMSGTITFM	119				
Db	67	ceelvriqtlnhmealgwdipgsflvgngkvvyegswlhvgahtyg-ynsrsglvafv	125				
Qy	120	GNEMDRVPAKRALRAALMLLECGVSRGFLRSNVEYKGRDVGQSTLPSGDQLYQVIOQSEH	179				
Db	126	gnfntdepsgamlealrllrcgverghlagdryvrvaqrliasesprklynqirrrwe	185				
Qy	180	YRE	182				

RESULT 15

Search completed: December 17, 2001, 07:49:31
Job time: 54 sec

OM of: US-09-462-625-2 to: GenEmbl.* out_format : pfs
 Date: Dec 17, 2001 11:22 AM
 About: Results were produced by the GenCore software, version 4.5,
 Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlp
 -O=/cgn2.1/USPTO.spool/US09462625/runat_17122001_074947_26111/app_query.fasta_1.241
 -DB=GenEmbl -QFMT=fastap -SUFFIX=p2n.rge -GAPOP=12.000
 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
 -QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
 -DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
 -TRANS=human40.cdi -LIST=45 -DOCALLIGN=200 -THR_SCORE=pct
 -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
 -NORM=ext -MINLEN=0 -MAXLEN=2000000000
 -USER=US09462625 -CGN1_1_0 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPXY
 -WAIT -THREADS=1

Search information block:

Query: US-09-462-625-2
 Query length: 182
 Database: GenEmbl.*
 Database sequences: 1472140
 Database length: 34134837
 Search time (sec): 1561.970000

score_list:

Sequence	Strid	Orig	ZScore	PScore	Len	Documentation
gb_pat:AR124884	+	996.00	1801.02	4.8e-22	549	AR124884 Sequence 1 from patent
gb_ro:AF193843	+	996.00	1799.30	6.0e-22	669	AF193843 Mus musculus TAG7-like
gb_ro:AF076482	+	996.00	1799.16	6.2e-22	680	AF076482 Mus musculus peptidog
gb_ro:BC005582	+	996.00	1798.74	6.5e-22	713	BC005582 Mus musculus, peptidog
gb_ro:NMNRNATM51	+	950.50	1716.51	2.5e-27	678	X86374 M.musculus mRNA for TAG7
gb_ro:AF154114	+	838.50	1513.65	4.9e-26	630	AF154114 Rattus norvegicus pept
gb_om:CDR154116	+	713.00	1284.71	2.8e-23	700	AF131676 Camelus dromedarius mR
gb_pr:AF076483	+	692.00	1246.68	3.6e-21	690	AF076483 Homo sapiens peptidog
gb_pr:AF242517	+	692.00	1246.26	3.9e-21	724	AF242517 Homo sapiens peptidog
gb_ro:MMTAG7	+	450.50	792.19	7.5e-36	4205	Y12088 Mus musculus tag7 gene,
gb_pat:AX1119918	+	406.00	722.91	5.6e-32	1110	AX119918 Sequence 7 from Paten
gb_pat:AX1119915	+	406.00	722.77	5.6e-32	1128	AX119915 Sequence 4 from Paten
gb_pr:AF242518	+	393.50	695.72	1.8e-20	1858	AF242518 Homo sapiens hypothet
gb_in:AC007085	+	384.00	639.38	2.5e-27	168667	AC007085 Drosophila melanoga
gb_in:AC007303	+	384.00	639.27	2.5e-27	169509	AC007303 Drosophila melanoga
gb_htg:AC020451	+	384.00	639.07	2.5e-27	172838	AC020451 Drosophila melanoga
gb_in:AE003836	+	384.00	635.75	3.9e-27	253273	AE003836 Drosophila melanoga
gb_in:AF207542	+	381.00	682.11	1.0e-29	652	AF207542 Drosophila melanogast
gb_in:AF207541	+	381.00	680.31	1.0e-28	717	AF207541 Drosophila melanogast
gb_in:AF076481	+	362.50	648.26	7.8e-28	670	AF076481 Trichoplusia ni peptid
gb_htg:AC014935	+	362.00	599.77	3.9e-25	159930	AC014935 Drosophila melanoga
gb_in:AC006933	+	362.00	598.46	4.7e-25	186002	AC006933 Drosophila melanoga
gb_in:AC003526	+	362.00	594.88	7.4e-25	280887	AC003526 Drosophila melanoga
gb_pr:AC007785	+	353.50	583.98	3.0e-24	166500	AC007785 Homo sapiens chromo
gb_pat:AX119920	+	344.00	610.28	1.0e-25	1107	AX119920 Sequence 9 from Paten
gb_pat:AX119917	+	344.00	610.14	1.0e-25	1125	AX119917 Sequence 6 from Paten
gb_in:AB017519	+	335.50	599.16	4.4e-25	673	AB017519 Bombyx mori mRNA for E
gb_in:AF313391	+	334.00	590.23	1.3e-24	1375	AF313391 Drosophila melanogast
gb_pat:E16479	+	325.50	582.13	3.8e-24	590	E16479 Bombyx mori mRNA for pef
gb_pat:E16480	+	325.50	580.01	5.0e-24	753	E16480 Bombyx mori mRNA for pef
gb_htg:AC018100	+	322.00	536.24	1.4e-21	55863	AC018100 Drosophila melanogast
gb_in:AE003558	+	322.00	522.43	8.0e-21	273744	AE003558 Drosophila melanoga
gb_in:AB016249	+	321.50	572.96	1.2e-23	735	AB016249 Bombyx mori mRNA for E
gb_in:AF207538	+	319.00	564.68	3.5e-23	1130	AF207538 Drosophila melanogast
gb_in:AF207537	+	319.00	564.46	3.6e-23	1158	AF207537 Drosophila melanogast
gb_in:AF035445	+	316.00	561.90	5.1e-23	831	AF035445 Calpodex ethilis cutic
gb_htg:AC013835	+	316.00	530.39	2.9e-21	31218	AC013835 Drosophila melanogast
gb_in:AE003552	+	316.00	511.12	3.4e-20	286784	AE003552 Drosophila melanoga
gb_pat:E16478	+	313.50	561.41	5.4e-23	521	E16478 Bombyx mori mRNA for pef
gb_htg:AC014935	+	306.50	498.93	1.6e-19	159930	AC014935 Drosophila melanoga

gb_in:AC006933 - 306.50 497.62 1.9e-19 185002 ! AC006933 Drosophila melan
 gb_in:AE003526 + 306.50 494.04 3.1e-19 280887 ! AE003526 Drosophila melan
 gb_in:AB017520 + 298.00 531.71 2.4e-21 622 ! AB017520 Bombyx mori mRNA fo
 gb_ro:AF149837 + 297.00 520.65 1.0e-20 1803 ! AF149837 Mus musculus TAGL-
 gb_htg:AC010028 + 294.50 480.61 1.7e-18 107216 ! AC010028 Drosophila melan
 seq_name: gb_pat:AR124884

seq documentation block:
 LOCUS AR124884 549 bp DNA PAT 16-MAY-2001
 DEFINITION Sequence 1 from patent US 6172211.
 ACCESSION AR124884
 VERSION AR124884.1 GI:14110245
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unknown.
 UNCLASSIFIED
 REFERENCE 1 (bases 1 to 549)
 AUTHORS Georgiev,G.P., Kiselev,S.L., Prokhorchouk,E.B. and Ostermann,E.
 TITLE Nucleic acid encoding tag7 polypeptide
 JOURNAL Patent: US 6172211-A 1 09-JAN-2001;
 FEATURES
 Location/Qualifiers
 source 1..549
 BASE COUNT 117 a 161 c 151 g 120 t
 ORIGIN

alignment_scores:

Quality: 996.00 Length: 182
 Ratio: 5.473 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-462-625-2 x AR124884 ..
 Align seg 1/1 to: AR124884 from: 1 to: 549

1 MetLeuPheAlaCysAlaLeuLeuAlaLeuLeuGlyLeuAlaThrSerCy 17
 1 ATGTTGTTGGCTGTGCTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTG 50

17 sSerPheIleValProArgSerGluTrpArgAlaLeuProSerGluCys 34
 1 CAGTTTCATGTCGCCGACGACGAGTGGAGGCGCTGCTGCTGCTGCTGCTG 100

34 erSerArgLeuGlyHisProValargytrValValIleSerHisThrala 50
 101 CTAGCCGCTGGGCGACCCACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 150

51 GlySerPheCysAsnSerProAspSerCysGluGlnGlnAlaArgAsnVa 67
 151 GGCAGCTCTGTCACACAGCCGCGAGTCTCTGCTGCTGCTGCTGCTGCTGCTG 200

67 lGlnHisTyrrHisLysAsnGluLeuGlyTrpCysAspValalaTyrrAsn 84
 201 GCAGCATTCACACAGAGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 250

84 heLeulleGlyGluAspGlyHisValTyrrGluGlyArgGlyTrpAsnIle 100
 251 TCCCTATTGGAGAGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 300

101 LysGlyAspHisThrGlyProIleTrpAsnProMetSerIleGlyIleTh 117
 301 AAGGGTGACACACAGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 350

117 rPheMeGlyAsnPheMetAspArgValProAlaLysArgAlaLeuArgA 134
 351 CTTATGGGGAACCTTCATGGAGCGGGTACCGCGAAAGCGGGCCCTCCCGTG 400

134 laAlaLeuAsnLeuLeuGluCysGlyValSerArgGlyPheLeuArgSer 150
 401 CTGCGCCATAAATCTTCTGGAATGTTGGGCTGCTGCGGGCTTCTCGAGATCC 450

```

151 AsnTyrGluValLysGlyHisArgAspValGlnSerThrLeuSerProG1 167
|||||
451 AACTATGAAGTCAAGGACACCGGGATGTGCAAGACACTCTCTCTCCAGG 500
|||||
167 YAspGlnLeuTyrGlnValLleGlnSerTyrGluHisTyrArgGlu 182
|||||
501 TGACCAACTATATCAGTCTATCCAAAGCTGGGAACACTACCGAGAG 546
|||||

seq_name: gb_ro:AF193843

seq_documentation_block: 669 bp mRNA ROD 07-NOV-1999
LOCUS AF193843
DEFINITION Mus musculus TAG7-like protein mRNA, complete cds.
ACCESSION AF193843
VERSION AF193843.1 GI:6273360
KEYWORDS
SOURCE house mouse.
ORGANISM Mus musculus
Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE 1 (bases 1 to 669)
AUTHORS Slayton,W.B., Rigaa,A., Hancock,J.D., Zaugg,J.K., Le,T.V.,
Trautman,M.S., Spangrude,G.J., Carroll,W.L. and Schibler,K.R.
TITLE Granulocyte-colony stimulating factor up-regulates expression of
murine tag7 during myeloid differentiation
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 669)
AUTHORS Slayton,W.B., Rigaa,A., Hancock,J.D., Zaugg,J.K., Le,T.V.,
Trautman,M.S., Spangrude,G.J., Carroll,W.L. and Schibler,K.R.
TITLE Direct Submission
JOURNAL Submitted (11-OCT-1999) Pediatrics, University of Utah, 50 North
Medical Drive, Room 2A126, Salt Lake City, UT 84132, USA
FEATURES
Source
1..669
/organism="Mus musculus"
/db_xref="taxon:10090"
/cell_line="32D cl3"
/cell_type="myeloblast"
25..573
/codon_start=1
/product="TAG7-like protein"
/protein_id="AA06335.1"
/db_xref="GI:6273361"
translation="MLFACALLALLGLATSCSFIVPRSEWRALPSECSRRLGHPVRV
VISHTAGSFNCSPDCSQARNVQHYHKNELGWCDVAYNFLIGEDGHVIEGRGNWIKG
DHTGPIWNPMSIGITFMGNFMDRVPAPKRALRAALLLECGVSRGFLRSNVEVKGHRD
QSTLSPGDLYQVTSWEHYRE"
BASE COUNT 144 a 212 c 173 g 140 t
ORIGIN

alignment_scores:
Quality: 996.00 Length: 182
Ratio: 5.473 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-462-625-2 x AF193843 ..
Align seg 1/1 to: AF193843 from: 1 to: 669

1 MetLeuPheAlaCysAlaLeuLeuAlaLeuLeuGlyLeuAlaThrSerCy 17
|||||
25 ATGTTGTTGCTGCTCTCTCTCTGCTCTCTGCTGCTGCTGCTGCTGCTG 74
|||||
17 sSerPheIleValProArgSerGluTyrArgAlaLeuProSerGluCys 34
|||||
75 CAGTTTCATCGTCCCGCAGTGTAGTGGAGGCGCTCCATCCAGTGCT 124
|||||
34 erSerArgLeuGlyHisProValArgTyrValLleSerHisThraLa 50
|||||
125 CTAGCCGCTGGGCGACCCAGCTTCGTCAGTGGTGGTATCTCACACAGCC 174
|||||

```

```

51 GlySerPheCysAsnSerProAspSerCysGluGlnGlnAlaArgAsnVa 67
|||||
175 GGCAGCTTCTGCACACAGCCCGGACTCTGTGACACAGCAGCGCCGCAATGT 224
|||||
67 lGlnHisTyrHisLysAsnGluLeuGlyTyrCysAspValAlaTyrAsnP 84
|||||
225 GCAGCATTACCACAAGAAATGAGTGGCTGTGCGATGTAGCCTTACAACT 274
|||||
84 heLeuIleGlyGluAspGlyHisValTyrGluGlyArgGlyTyrAsnIle 100
|||||
275 TCCTTATTGGAGAGGACGGTCATGCTATGAAGCCGAGGCTGGGAACATC 324
|||||
101 LysGlyAspHisThrGlyProIleTyrAsnProMetSerIleGlyIleTh 117
|||||
325 AAGGTGACCACACAGCGCCCATCTGGAAATCCCATGCTATTGGCATCAC 374
|||||
117 rPheMetGlyAsnPheMetAspArgValProAlaLysArgAlaLeuArgA 134
|||||
375 CTTTCATGGGGAACCTTCATGGACCGGTACCCGCAAGCGGCCCTCCGTG 424
|||||
134 laAlaLeuAsnLeuLeuGluCysGlyValSerArgGlyPheLeuArgSer 150
|||||
425 CTGCCCTAAATCTTCTTGAATGTGGGTGCTCGGGGCTTCTCTGAGATCC 474
|||||
151 AsnTyrGluValLysGlyHisArgAspValGlnSerThrLeuSerProG1 167
|||||
475 AACTATGAAGTCAAGGACACCGGGATGTGCAAGACACTCTCTCTCCAGG 524
|||||
167 YAspGlnLeuTyrGlnValLleGlnSerTyrGluHisTyrArgGlu 182
|||||
525 TGACCAACTATATCAGTCTATCCAAAGCTGGGAACACTACCGAGAG 570
|||||

seq_name: gb_ro:AF076482

seq_documentation_block: 680 bp mRNA ROD 15-AUG-1998
LOCUS AF076482
DEFINITION Mus musculus peptidoglycan recognition protein precursor (Pgrp)
mRNA, complete cds.
ACCESSION AF076482
VERSION AF076482.1 GI:3342530
KEYWORDS house mouse.
SOURCE Mus musculus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE 1 (bases 1 to 680)
AUTHORS Kang,D., Liu,G., Lundstrom,A., Gelius,E. and Steiner,H.
TITLE A peptidoglycan recognition protein in innate immunity conserved
from insects to humans
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 95 (17), 10078-10082 (1998)
MEDLINE 98374308
REFERENCE 2 (bases 1 to 680)
AUTHORS Kang,D., Liu,G., Lundstrom,A., Gelius,E. and Steiner,H.
TITLE Direct Submission
JOURNAL Submitted (08-JUL-1998) Microbiology, Stockholm University,
Stockholm S-106 91, Sweden
FEATURES
Location/Qualifiers
1..680
/organism="Mus musculus"
/db_xref="taxon:10090"
/tissue_type="spleen"
1..680
/gene="Pgrp"
36..584
/gene="Pgrp"
/note="Pgrp"
/codon_start=1
/product="peptidoglycan recognition protein precursor"
/protein_id="AAC31821.1"
/db_xref="GI:3342531"
/translation="MLFACALLALLGLATSCSFIVPRSEWRALPSECSRRLGHPVRV
VISHTAGSFNCSPDCSQARNVQHYHKNELGWCDVAYNFLIGEDGHVIEGRGNWIKG

```



```

34  erSerArgLeuGlyHisProValArgTyrValValIleSerHisThrAla 50
147 CTAGCGCGCTGGGGCACCAGTTCGCTAGTGGTGATCTCACACAGCC 196

51  GlySerPheCysAsnSerProAspSerCysGluGlnGlnAlaArgAsnVa 67
197 GGCACCTTCTGACACGCCCGGACTCCTGTGTGACACAGCAGGCCCGCAATGT 246

67  lGlnHisTyrHisLysAsnGluLeuGlyTrpCysAspValAlaTyrAsnP 84
247 GCAGCATTAACCAAGAATGAGCTGGCTGGTGCATGTAGCCTACAAC 296

84  heLeuIleGlyAspGlyHisValTyrGluGlyArgGlyTrpAsnIle 100
297 TCCTTATTGGAGAGCGGTCTATGTGAAGCGCGAGGCTGGAACATC 346

101 LysGlyAspHisThrGlyProIleTrpAsnPrometSerIleGlyIleTh 117
347 AAGGTGACCAACAGGGCCCATCTGGAATCCATGTCTATTGGCATCAC 396

117 rPheMetGlyAsnPheMetAspArgValProAlaLysArgAlaLeuArgA 134
397 CTTTCATGGGGAACCTTCATGACCGGTACCCGCAAGCGGGCCCTCCGTG 446

134 laAlaLeuAsnLeuLeuGlyCysGlyValSerArgGlyPheLeuArgSer 150
447 CTGCCCTAAATCTTCTGGAATGTGGGGTGTCTCGGGGCTCTTGATGCC 496

151 AsnTyrGluValLysGlyHisArgAspValGlnSerThrLeuSerProGl 167
497 AACTATGAAGTCAAGGACACCGGGATGTGCAAGACACTCTCTCTCCAGG 546

167 yAspGlnLeuTyrGlnValIleGlnSerTrpGluHisTyrArgGlu 182
547 TGACCAACTCTATCAGGTCTATCAAGTCTCAAGTGGGAACACTACCGGAG 592

seq_name: gb_ro:MMRNATMS1

seq_documentation_block: 678 bp mRNA ROD 30-JUL-1998
LOCUS MMRNATMS1
DEFINITION M.musculus mRNA for TAG7 protein.
ACCESSION X86374
VERSION X86374.1 GI:887524
KEYWORDS tms1 gene.
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE
AUTHORS Kustikova,O.S., Kiselev,S.L., Borodulina,O.R., Senin,V.M.,
Afanas'eva,A.V. and Kabishev,A.A.
TITLE Cloning of the tag7 gene expressed in metastatic mouse tumors
JOURNAL Genetika 32 (5), 621-628 (1996)
MEDLINE 96321104
REFERENCE
AUTHORS Kustikova,O.S.
TITLE Direct Submission
JOURNAL Submitted (18-APR-1995) O.S. Kustikova, Inst of Gene Biology
Russian Academy, of Science, B334, Vavilov Street 34/5, Moscow,
RUSSIA
REFERENCE
AUTHORS Kiselev,S.L., Kustikova,O.S., Korobko,E.V., Prokhortchouk,E.B.,
Kabishev,A.A., Lukanidin,E.M. and Georgiev,G.P.
TITLE Molecular cloning and characterization of the mouse tag7 gene
encoding a novel cytokine
JOURNAL J. Biol. Chem. 273 (29), 18633-18639 (1998)
MEDLINE 98325081
COMMENT On Jul 6, 1995 this sequence version replaced gi:785010.
FEATURES
LOCATION/Qualifiers
1..678
/organism="Mus musculus"
/strain="A/Sn"
source

/db_xref="taxon:10090"
/tissue_type="adenocarcinoma"
/dev_stage="adult"
36..491
/gene="tag7"
36..491
/gene="tag7"
/codon_start=1
/protein_id="CAA60133.1"
/db_xref="GI:887525"
/db_xref="MGI:107415"
/translat="SPTREMBL:O62185"
VISTAGSFNCSPDSCQEQARNVQHYHKNELGWCFYVNFLLIGEDGHVYEGRGWNIKG
DHTGPINPMSIGITFMGNFMDRVKAGPPCCPKSGIWSGLPEIQL"
BASE COUNT 146 a 216 c 173 g 143 t
ORIGIN

alignment_scores:
Quality: 950.50 Length: 183
Ratio: 5.310 Gaps: 2
Percent Similarity: 97.814 Percent Identity: 97.814

alignment_block:
US-09-462-625-2 x MMRNATMS1 ..
Align seg 1/1 to: MMRNATMS1 from: 1 to: 678

1 MetLeuPheAlaCysAlaLeuLeuAlaLeuLeuGlyLeuAlaThrSerCy 17
|||||
36 ATGTTGTTTGCCTGTCTCTCTGCTTGGCTTGGCTTGGCAACCTCTCTG 85
|||||

17 sSerPheIleValProArgSerGluTrpArgAlaLeuProSerGluCysS 34
|||||
86 CAGTTTCATCGTGGCCCGCAGTAGTGAGGGCCCTGCCATCCGAGTGT 135
|||||

34 erSerArgLeuGlyHisProValArgTyrValValIleSerHisThrAla 50
|||||
136 CTAGCGCGCTGGGGCACCAGTTCGCTAGTGGTGATCTCACACAGCC 185
|||||

51 GlySerPheCysAsnSerProAspSerCysGluGlnGlnAlaArgAsnVa 67
|||||
186 GCGAGCTTCTGCAACAGCCCGGACTCTCTGTGAACAGCAGGCCCGCAATGT 235
|||||

67 lGlnHisTyrHisLysAsnGluLeuGlyTrpCysAspValAlaTyrAsnP 84
|||||
236 GCAGCATTAACCAAGAATGAGCTGGCTGGTGCATGTAGCCTACAAC 285
|||||

84 heLeuIleGlyGluAspGlyHisValTyrGluGlyArgGlyTrpAsnIle 100
|||||
286 TCCTTATTGGAGAGCGGTCTATGTGAAGCGCGAGGCTGGAACATC 335
|||||

101 LysGlyAspHisThrGlyProIleTrpAsnPrometSerIleGlyIleTh 117
|||||
336 AAGGTGACCAACAGGGCCCATCTGGAATCCATGTCTATTGGCATCAC 385
|||||

117 rPheMetGlyAsnPheMetAspArgValProAlaLysArgAlaLeuArg 133
|||||
386 CTTTCATGGGGAACCTTCATGACCGGTAC...GCAAGCGGGCCCTCCGT 432
|||||

134 AlaAlaLeuAsnLeuLeuGlyCysGlyValSerArgGlyPheLeuArgSe 150
|||||
433 GCTGCCCTAAATCTTCTGGAATCTGGGGTGTCTCGGGGCTTCTGATC 482
|||||

150 rAsnTyrGluValLysGlyHisArgAspValGlnSerThrLeuSerProG 167
|||||
483 CAACATATGAAGTCAAGGACACCGGGATGTGCAAGCACTCTCTCTCCAG 532
|||||

167 lyAspGlnLeuTyrGlnValIleGlnSerTrpGluHisTyrArgGlu 182
|||||
533 GTGACCAACTCTATCAGGTCTATCAAGTGGGAACACTACCGGAGAG 579
|||||

```

alignment_scores:


```

BASE COUNT      249 a      319 c      295 g      247 t
ORIGIN

alignment_scores:
  Quality: 406.00      Length: 169
  Ratio: 3.355      Gaps: 2
  Percent Similarity: 71.598      Percent Identity: 43.787

alignment_block:
US-09-462-625-2 x AX119918      ..

Align seg 1/1 to: AX119918 from: 1 to: 1110

13  LeuAlaThrSerCysSerPheIleValProArgSerGluTrpArgAlaLe 29
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
604  CTGAAGAAGGCTTGCCCGGGCTGTCCACGGTCTGTGGGGAGCCAG 653
29  uProSerGluCysSerSerArgLeuGlyHisProValArgTyrValValI 46
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
654  GGAGAGCCCACTGT...CCGAGATCACTCTCCAGGAGATATGCATCA 700
46  leSerHisThrAlaGlySerPheCysAsnSerProAspSerCysGluGln 62
||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
701  TTATCCACACTCCGGGAGGACTGCACACATTTCTGATGAGTGCGCCGTG 750
63  GlnAlaArgAsnValGlnHisTyrHisLysAsnGluLeuGlyTrpCysAs 79
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
751  CTGGTCGGGAGATCCAGCTTTCTACATAGACAGGCTCAAGTCATGCGA 800
79  pValAlaTyrAsnPheLeuIleGlyAspGlyHisValTyrGluGlyA 96
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
801  CATTGGTTATACTCTCGTGGCGCAGGATGGCGCAATTTATGAAGGG 850
96  rgGlyTrpAsnIleLysGlyAspHisThrGlyProIleTrpAsnPromet 112
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
851  TGGGCTTGAATCTCCAGGCTCCTCCACC...CCTGGCTACGATGACATT 897
113  SerIleGlyIleThrPheMetGlyAsnPheMetAspArgValProAlaLy 129
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
898  GCCTCGGCATTACCTTCTATGGGCACCTTCACAGGTATACCCACCAATGC 947
129  sArgAlaLeuArgAlaAlaLeuAsnLeuLeuGlyValSerArgG 146
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
948  TGCAGCACTAGAGGACGCCAAGACCTGATCCAGTGTGCCATGGTCAAA 997
146  lyPheLeuArgSerAsnTyrGluValLysGlyHisArgaspValGlnSer 162
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
998  GGTACCTGACTCCCAACTACCTCTGGTGGGCCACAGTATGTGGCCCGA 1047
163  ThrLeuSerProGlyAspGlnLeuTyrGlnValIleGlnSerTrpGluHI 179
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1048  ACCTTGTCCTCTGGGAGGCTTTGTACACATCATCAGCACCTGGCCTCA 1097

179  sTyrArg 181

```

16 SerCysSerPheIleValProArgSerGluTrpArgAlaLeuProSerG1 32
:::|||||
700 GCTTGCCCCCGCATGTGCCACGGTCTGTGTGGGGAGCC.....AGGGA 743

16 SerCysSerPheIleValProArgSerGluTrpArgAlaLeuProSerG1 32
:::|||||
700 GCTTGCCCCCGCATGTGCCACGGTCTGTGTGGGGAGCC.....AGGGA 743

32 ucysSer_SerArgLeuGlyHisProValArgTyrValValIleSerHis 48
 744 GACCACTGTCAGGATGACTCTCCAGCGAAGTATGCGATCATATCCAC 793
 49 ThrAlaGlySerPheCysAsnSerProAspSerCysGluGlnAlaAr 65
 794 ACTGCGGGAGGACCTGCACATCTTCTGATGAGTGCCTGCTGCTCG 843
 65 gAsnValGlnHisTyrHisLysAsnGluLeuGlyTyrCysAspValAla 82
 844 GGACATCCAGTCTTCTACATAGACAGGCTCAAGTCATCGACATGGTT 893
 82 TyrAsnPhelLeuGlyGluAspGlyHisValTyrGluGlyArgGlyTyr 98
 894 ATAACCTCTGTGGCCAGAGTGCCTATATGAGGGTGGGCTGG 943
 99 AsnIleLysGlyAspHisThrGlyProIleTyrAsnProMetSerIleG 115
 944 AATGTCACAGGCTCTCCACC...CTGGCTACGATGACATGCCCTGG 990
 115 yIleThrPheMetGlyAsnPhemMetAspArgValProAlaLysArgAla 132
 991 CATTACCTTCATGGCCACCTTCACAGGTATACACCAATGCTGCAGCAC 1040
 132 euArgAlaAlaLeuAsnLeuLeuGlyGlyValSerArgGlyPheLeu 148
 1041 TAGAGGCACCCCAAGACCTTCAGTCCAGTGCCTGATGTCGAAGGGTAC 1090
 149 ArgSerAsnTyrGluValLysGlyHisArgAspValGlnSerThrLeu 165
 1091 ACTCCCACTACCTGCTGTGGCCACAGTGTGGCCCGAACCTTGTC 1140
 165 rProGlyAspGlnLeuTyrGlnValIleGlnSerTrpGluHisTyrArg 181
 1141 TCCTGGGAGGCTTGTACAGATCATCAGCACCTGGCTCATTTCAAA 1189
 seq_name: gb_in:AC007085

seq_documentation_block:
 LOCUS AC007085 166867 bp DNA INV 17-MAR-2001
 DEFINITION Drosophila melanogaster, chromosome 2R, region 44D2-44E1, BAC clone
 BACR21H10, complete sequence.
 ACCESSION AC007085
 VERSION AC007085.7 GI:13374639
 KEYWORDS HYG.
 SOURCE fruit fly.
 ORGANISM Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
 Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 1 (bases 1 to 166867)
 Celniker,S.E., Adams,M.D., Kronmiller,B., Tyler,D., Wan,K.H.,
 Holt,R.A., Evans,C.A., Gocayne,J.D., Amanatides,P.G., Brandon,R.C.,
 Rogers,Y., An,H., Baldwin,D., Banazon,J., Beeson,K.Y., Busam,D.A.,
 Carlson,J.W., Center,A., Champs,M., Davenport,L.B., Dietz,S.M.,
 Dodson,K., Dorsett,V., Doup,E.E., Doyle,C., Dresnek,D., Farfan,D.,
 Ferreira,S., Frise,E., Galle,R.F., Garg,N.S., George,R.A.,
 Gonzalez,M., Hock,J., Hoskins,R.A., Hostin,D., Howland,T.J.,
 Ibegwam,C., Jalali,M., Kruse,D., Li,P., Mattei,B., Moshrefi,A.,
 McIntosh,T.C., Moy,M., Murphy,B., Nelson,C., Nelson,K.A., Nunoo,J.,
 Pacleb,J., Paragas,V., Park,S., Patel,S., Pfeiffer,B.,
 Phouanavong,S., Pittman,G.S., Puri,V., Richards,S., Scheeler,F.,
 Stapleton,M., Strong,R., Svirskas,R., Tector,C., Williams,S.M.,
 Zaveri,J.S., Smith,H.O., Rubin,G.M. and Venter,J.C.
 Sequencing of Drosophila chromosome 2R, region 44D2-44E1
 Unpublished
 2 (bases 1 to 166867)
 Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
 Butenhoff,C., Champs,M., Chavez,C., Chew,M., Ciesiolka,L.,
 Doyle,C.M., Farfan,D.E., Galle,R.A., George,R.A., Harris,N.L.,
 Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
 Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,
 Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,

Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E.,
 Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and
 Rubin,G.M.
 Direct Submission
 Submitted (16-MAR-1999) Drosophila Genome Center, Lawrence Berkeley
 Laboratory, MS 64-121, Berkeley, CA 94720, USA
 On Mar 17, 2001 this sequence version replaced gi:5670628.
 Sequence submitted by:
 Berkeley Drosophila Genome Project
 Lawrence Berkeley National Laboratory, MS 64-121,
 Berkeley, CA 94720
 This sequence was assembled using end sequences from a whole genome
 shotgun and from subclones of this BAC and its neighboring clones.
 For further information about this sequence, including its location
 and relationship to other sequences, please visit our sequence
 archive web site (<http://www.fruitfly.org/sequence/>) or send email
 to bdg@fruitfly.berkeley.edu.
 FEATURES
 Location/Qualifiers
 1..166867
 /organism="Drosophila melanogaster"
 /strain="y; cn bw sp"
 /db_xref="taxon:7227"
 /chromosome="2R"
 /map="44D2-44E1"
 /clone="BACR21H10 (D580)"
 /clone_lib="RFCI-98 (Roswell Park Cancer Institute
 Drosophila melanogaster BAC library, partial EcoRI in
 pBACe3.6)"
 BASE COUNT 46731 a 36967 c 36753 g 46416 t
 ORIGIN
 alignment_scores:
 Quality: 384.00 Length: 181
 Ratio: 2.931 Gaps: 4
 Percent Similarity: 72.376 Percent Identity: 42.541
 alignment_block:
 US-09-462-625-2 x AC007085 ..
 Align seg 1/1 to: AC007085 from: 1 to: 166867
 1 MetLeuPheAlaCysAlaLeuLeuAlaLeuLeuGlyLeuAlaThrSerCy 17
 159245 GFACTCTTC...TGCGCC...CAAGCGGTCTCGCGTGACC..... 159280
 17 sSerPheIleValProArgSerGluTrpArgAlaLeuProSerGluCys 34
 159281ATCATCTCCAAGTCGGAGTGGGGCGCGTTCGCCACGAGCA 159323
 34 erSerArgLeuGlyHisProValArgTyrValValIleSerHisThrAla 50
 159324 AGACCTCGCTGGCCAACTACTGAGTACCGCGTGTATCCACACACCGCT 159373
 51 GlySerPheCysAsnSerProAspSerCysGluGlnGlnAlaArgAsnVa 67
 159374 GGAATCTACTGACAGCAGCAGCGCGCTGTCATCACAGCTGCAGACAT 159423
 67 lGlnHisTyrHisLysAsnGluLeuGlyTyrCysAspValAlaTyrAsn 84
 159424 CCAGGCTACACATGACATGCCCTGGCTGGCGGATATCGGCTACAACT 159473
 84 heLeuIleGlyGluAspGlyHisValTyrGluGlyArgGlyTyrAsnIle 100
 159474 TCCTGATCGCGGAGAGCGCAACGTGTACAGGGTTCGCGCTGGAACGTT 159523
 101 LysGlyAspHisThrGlyProIleTyrAsnProMetSerIleGlyIleTh 117
 159524 ATGGGTCCTCAC...GCCACTACTGGAACCTCCAAAGTCTATCGGCATCTC 159570
 117 rPheMetGlyAsnPhemMetAspArgValProAlaLysArgAlaLeuArg 134
 159571 CTTCTGGGCACTACAATACCAACACACCCCTCACCTCTGCTCAGATCACCG 159620

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 17, 2001, 07:48:00 ; Search time 1560.95 Seconds
(without alignments)
5802.205 Million cell updates/sec

Title: US-09-462-625-1

Perfect score: 549

Sequence: 1 atgttttgcctgtgctct.....gggaacactaccagagatga 549

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*

1: gb.ba.*

2: gb.htg.*

3: gb.in.*

4: gb.ov.*

5: gb.pat.*

6: gb.ph.*

7: gb.pl.*

8: gb.pr.*

9: gb.ro.*

10: gb.sts.*

11: gb.sy.*

12: gb.un.*

13: gb.vi.*

14: gb.ba.*

15: em.fun.*

16: em.hum.*

17: em.in.*

18: em.om.*

19: em.or.*

20: em.ov.*

21: em.pat.*

22: em.ph.*

23: em.pl.*

24: em.ro.*

25: em.sts.*

26: em.sy.*

27: em.un.*

28: em.vi.*

29: em.htgo_hum.*

30: em.htgo_inv.*

31: em.htgo_rod.*

32: em.htg_hum.*

33: em.htg_inv.*

34: em.htg_rod.*

35: em.htg_other.*

36: em.htg_other.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

8

Result No.	Score	Query Match	Length	DB	ID	Description
1	549	100.0	549	6	ARI24884	ARI24884 Sequence
2	549	100.0	669	10	AF193843	AF193843 Mus muscu
3	547.4	99.7	680	10	AF076482	AF076482 Mus muscu
4	547.4	99.7	713	10	BC005582	BC005582 Mus muscu
5	533.4	97.2	678	10	MMRNATMS1	X86374 M.musculus
6	452.8	82.5	630	10	AF154114	AF154114 Rattus no
7	316.6	57.7	690	9	AF076483	AF076483 Homo sapi
8	316.6	57.7	724	9	AF242517	AF242517 Homo sapi
9	307.8	56.1	700	4	CDRI31676	AF131676 Camelus d
10	249.4	45.4	4205	10	NMTAG7	Y12088 Mus musculus
11	131.6	24.0	166867	3	AC007085	AC007085 Drosophill
12	131.6	24.0	169509	3	AC007303	AC007303 Drosophill
13	131.6	24.0	172838	2	AC020451	AC020451 Drosophill
14	131.6	24.0	253273	3	AE003836	AE003836 Drosophill
15	127	23.1	166500	9	AC007785	AC007785 Homo sapi
16	126.8	23.1	1107	6	AX119920	AX119920 Sequence
17	126.8	23.1	1125	6	AX119917	AX119917 Sequence
18	122.6	22.3	717	3	AF207541	AF207541 Drosophill
19	118.4	21.6	652	3	AF207542	AF207542 Drosophill
20	107	19.5	1110	6	AX119918	AX119918 Sequence
21	107	19.5	1128	6	AX119915	AX119915 Sequence
22	105.8	19.3	1858	9	AF242518	AF242518 Homo sapi
23	88.6	16.1	622	3	AB017520	AB017520 Bombyx mo
24	84.4	15.4	670	3	AF076481	AF076481 Trichoplu
25	79.4	14.5	1375	3	AF313391	AF313391 Drosophill
26	71	12.9	521	6	E16478	E16478 Bombyx mori
27	71	12.9	590	6	E16479	E16479 Bombyx mori
28	71	12.9	753	6	E16480	E16480 Bombyx mori
29	69.4	12.6	735	3	AB016249	AB016249 Bombyx mo
30	69	12.6	1130	3	AF207538	AF207538 Drosophill
31	69	12.6	1158	3	AF207537	AF207537 Drosophill
32	68	12.4	31218	2	AC013835	AC013835 Drosophill
33	68	12.4	159930	2	AC014935	AC014935 Drosophill
34	68	12.4	186002	3	AC006933	AC006933 Drosophill
35	68	12.4	280887	3	AE003526	AE003526 Drosophill
36	68	12.4	286784	3	AE003552	AE003552 Drosophill
37	67.8	12.3	831	3	AF035445	AF035445 Calpodes
38	67.6	12.3	55863	2	AC018100	AC018100 Drosophill
39	67.6	12.3	273744	3	AE003558	AE003558 Drosophill
40	64.8	11.8	107216	2	AC010028	AC010028 Drosophill
41	64.4	11.7	1716	10	AF149838	AF149838 Mus muscu
42	64.4	11.7	1803	10	AF149837	AF149837 Mus muscu
43	64.2	11.7	673	3	AB017519	AB017519 Bombyx mo
44	63.8	11.6	27655	2	AC019935	AC019935 Drosophill
45	63.8	11.6	185200	3	AC092401	AC092401 Drosophill

ALIGNMENTS

RESULT 1

ARI24884	ARI24884	549 bp	DNA	PAT	16-MAY-2001
LOCUS	Sequence 1 from patent US 6172211.				
DEFINITION	Sequence 1 from patent US 6172211.				
ACCESSION	ARI24884				
VERSION	ARI24884.1	GI:14110245			
KEYWORDS	Location/Qualifiers				
SOURCE	Unknown.				
ORGANISM	Unknown.				
REFERENCE	1 (bases 1 to 549)				
AUTHORS	Georgiev,G.P., Kiselev,S.L., Prokhorchouk,E.B. and Ostermann,E.				
TITLE	Nucleic acid encoding tag7 polypeptide				
JOURNAL	Patent: US 6172211-A 1 09-JAN-2001;				
FEATURES	Location/Qualifiers				
source	1..549				
BASE COUNT	117 a 161 c 151 g 120 t				
ORIGIN					

AUTHORS Kang, D., Liu, G., Lundstrom, A., Gelius, E. and Steiner, H.
 TITLE A peptidoglycan recognition protein in innate immunity conserved
 from insects to humans
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 95 (17), 10078-10082 (1998)
 MEDLINE 98374308
 REFERENCE 2 (bases 1 to 680)
 AUTHORS Kang, D., Liu, G., Lundstrom, A., Gelius, E. and Steiner, H.
 TITLE Direct Submission
 JOURNAL Submitted (08-JUL-1998) Microbiology, Stockholm University,
 Stockholm S-106 91, Sweden

FEATURES
 source Location/Qualifiers
 1..680
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /tissue_type="spleen"
 gene 1..680
 /gene="Pgrp"
 CDS 36..584
 /gene="Pgrp"
 /note="PGRP"
 /codon_start=1
 /product="peptidoglycan recognition protein precursor"
 /protein_id="AAC31821.1"
 /db_xref="GI:3342531"
 /translation="MLFACALLALLGLATSCSFIVPRSEWRALPSECSRLLGHVPRVY
 VISHTAGSFNCSPDSCQQARNVQHYHKNELGWCVDVAYNFIIGEDGHVYEGRWNIKG
 DHTGPIWNPMSIGITFMGNFMDRVPAPKRALRALNLLCEGSRGFLRSNVEYKGRD
 QSTLSPGDOLYQVIOQSWHYRE"
 sig_peptide 36..83
 /gene="Pgrp"
 mat_peptide 84..581
 /gene="Pgrp"
 /product="peptidoglycan recognition protein"
 BASE COUNT 146 a 216 c 174 g 144 t
 ORIGIN

Query Match 99.7%; Score 547.4; DB 10; Length 680;
 Best Local Similarity 99.8%; Pred. No. 2.8e-131;
 Matches 548; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 attgtatttcctgtgctctccttgcctcctcctggttggcgaacctcctcagtttcac 60
 DB |||||
 QY 61 gtgcccgcagtgagtgagggccctccatccagtgctctagccctcctgggaccca 120
 DB |||||
 QY 96 GTCCCCCGCAGTGAGTGAGGGGCCCTGCCATCCGAGTGCCTAGCCGCCCTGGGCAATCCA 155
 QY 121 gttcgtaagtgtgatctcacacagccggcagcttctgcaacagccggcagctcgt 180
 DB |||||
 QY 156 GTTCGCTAGCTGTGTATCTCACACAGCGCGAGCTTCTGCAACAGCCGAGCTCTGT 215
 QY 181 gaacacagggccgcaatgtgagcattaccacaagaatagctgggtggtgogatgta 240
 DB |||||
 QY 216 GAACACAGGCCCGCAATGTGACAGATTACCAAGAATGAGTGGCTGTGCGATGTA 275
 QY 241 gctcaaaccttcttattgagagggcgtcatgtctatgaagccgagcgtggaacatc 300
 DB |||||
 QY 276 GCTACAACTCTTATTGAGAGGAGCGTCAATGCTATGAAGCCGAGGCTGGAACATC 335
 QY 301 aagggtagcacacagggcccatctggaatccccatgctctattggcatcaacttcattggg 360
 DB |||||
 QY 336 AAGGGTGACACACAGGGCCCATCTGGAATCCCATGCTCTATTGGCATCACCTTCATGGGG 395
 QY 361 aacttcattgacgggtaccgcaagggccctcctcgtgctgctccctaaatctcttgaa 420
 DB |||||
 QY 396 AACTTCATGACCGGGTACCGCAAGCGGCCCTCCGTGCTGCTCCCTAAATCTTCTGGAA 455
 QY 421 tgtgggggtgtctcggggtcctcctgagatccaactatgaagtcaaaagacacccggatgtg 480
 DB |||||
 QY 456 TGTGGGGGTGTCTCGGGGCTCTCTGAGATCCAACTATGAAGTCAAGGACACACCGGGATGTG 515

QY 481 caaagcactctctccagtgaccacactctatcaggtcatccaaagctgggaacactac 540
 DB |||||
 QY 516 CAAAGCACTCTCTCCAGGTGACCAACTCTATCAGGTCTATCCAAAGCTGGGACACTAC 575
 DB |||||

QY 541 cagagtgta 549
 DB |||||

QY 576 CGAGAGTGA 584
 DB |||||

RESULT 4
 BC005582 713 bp mRNA ROD 12-JUL-2001
 LOCUS Mus musculus, peptidoglycan recognition protein, clone MGC:11430
 DEFINITION IMAGE:3969014, mRNA, complete cds.
 ACCESSION BC005582
 VERSION BC005582.1 GI:13542755
 KEYWORDS MGC.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 713)
 Strausberg, R.
 Direct Submission
 Submitted (02-APR-2001) National Institutes of Health, Mammalian
 Gene Collection (MGC), Cancer Genomics Office, National Cancer
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
 USA
 NIH-MGC Project URL: http://mgc.nci.nih.gov
 Contact: MGC help desk
 Email: cgabs-r@mail.nih.gov
 Tissue Procurement: Gilbert Smith, Ph.D.
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Baylor College of Medicine Human Genome
 Sequencing Center
 Center code: BCM-HGSC
 Web site: http://www.hgsc.bcm.tmc.edu/cdna/
 Callout: villalon@bcm.tmc.edu.
 Villalon, D.K., Luna, R.A., Hale, S.M., Hulyk, S., Lu, X., Garcia,
 A.M., Holloway, M., Telford, B., Hodgson, A., Bouck, J., Yu, W.,
 Muzny, D.N., Gibbs, R.A.
 Clone distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
 Series: IRAC Plate: 17 Row: j Column: 22
 This clone was selected for full length sequencing because it
 passed the following selection criteria: matched mRNA gi: 3342530.

FEATURES
 source Location/Qualifiers
 1..713
 /organism="Mus musculus"
 /db_xref="LocusID:21946"
 /db_xref="taxon:10090"
 /clone="MGC:11430 IMAGE:3969014"
 /tissue_type="Mammary tumor. Metallothionien-TGF alpha
 model. 10 month old virgin mouse. Taken by biopsy."
 /clone_lib="NCI CGAP_Mam1"
 /lab_host="DH10B"
 /note="Vector: pCMV-SPORT6"
 47..595
 /codon_start=1
 /product="peptidoglycan recognition protein"
 /protein_id="AAH05582.1"
 /db_xref="GI:13542756"
 CDS
 172 a 218 c 178 g 145 t
 BASE COUNT
 ORIGIN

Query Match 99.7%; Score 547.4; DB 10; Length 713;

	Query Match	97.2%;	Score 533.4;	DB 10;	Length 678;
	Best Local Similarity 99.3%;	Pred. No. 1.2e-127;			
	Matches 546;	Conservative 0;	Mismatches 1;	Indels 2;	Gaps 1;
Qy	1 atgttgttgctgctccttcggccctcctgggtctggcgaaacctcctcagtttcatc	60			
Dd	36 ATGTTTGTTCCTTGCTCTCCTTGCCCTCTGGGTCTGGCAACCTCTCGAGTTTCATC	95			
Qy	61 gtgcccccgcagtgaaggccctgcataccagatgctcttagccgctcgtgggaccca	120			
Dd	96 GTGCCCGCAGTAGTGGAGGGCCCTGCATCCGAGTGTCTAGCCGCTCGGGACACCA	155			
Qy	121 gttgcgtacgtggtgatctcacacagccgcagactcttgcacaagccccgagactcgt	180			
Dd	156 GTTCGCTACGTGGTGGATCTCACACAGCCGGCAGCTTCTGCAACACACCCGGACTCTCTGT	215			
Qy	181 qaacagcagggcccgaattgacgattaccacaagaatagctgggctggtgcgatgta	240			

Qy	241	gcctacaacttccttattggagagagcggctcatgtctatgaagccgagcgctggaacatc	300
Db <td>276</td> <td>GCCATACAACTTCCTTATTGGAGAGACGGTCATGCTATGAAGCCGAGGCTGGAAACATC</td> <td>335</td>	276	GCCATACAACTTCCTTATTGGAGAGACGGTCATGCTATGAAGCCGAGGCTGGAAACATC	335
Qy <td>301</td> <td>aagggtgaccacacagggcccacatcggaatcccattgtctattggcattcacaccttcattgggg</td> <td>360</td>	301	aagggtgaccacacagggcccacatcggaatcccattgtctattggcattcacaccttcattgggg	360
Db <td>336</td> <td>AAGGTGTACCCACACAGGGCCCATCTGGAATCCCATGCTATTGGCATCACCTTCATGGGG</td> <td>395</td>	336	AAGGTGTACCCACACAGGGCCCATCTGGAATCCCATGCTATTGGCATCACCTTCATGGGG	395
Qy <td>361</td> <td>aacttcattggaccgggtaccocgaaagcgggcccctcggctgcgcctataatcttctggaa</td> <td>420</td>	361	aacttcattggaccgggtaccocgaaagcgggcccctcggctgcgcctataatcttctggaa	420
Db <td>396</td> <td>AACCTCATGTGGACCGGGTA--CGCAAGCGGGCCCTCCGCTGCTGCCCTTAAATCTTCTGGAA</td> <td>453</td>	396	AACCTCATGTGGACCGGGTA--CGCAAGCGGGCCCTCCGCTGCTGCCCTTAAATCTTCTGGAA	453
Qy <td>421</td> <td>tgtgggggtgtctcggggcttctcagattccaactatgaagtc aaagacacccggatgtg</td> <td>480</td>	421	tgtgggggtgtctcggggcttctcagattccaactatgaagtc aaagacacccggatgtg	480
Db <td>454</td> <td>TCTGGGGTGTCTCGGGGCTTCCTGAGATCCAACTATGAAGTCNAAGGACACCGGGATGTG</td> <td>513</td>	454	TCTGGGGTGTCTCGGGGCTTCCTGAGATCCAACTATGAAGTCNAAGGACACCGGGATGTG	513
Qy <td>481</td> <td>caagcaactctctcgaagtgaccaactctatcagttcattcaaaagctgggaacactac</td> <td>540</td>	481	caagcaactctctcgaagtgaccaactctatcagttcattcaaaagctgggaacactac	540
Db <td>514</td> <td>CAAAAGCACTCTCTCCAGAGTGACCAACTCTATCAGGTCATCCAAAGCTGGGAACACTAC</td> <td>573</td>	514	CAAAAGCACTCTCTCCAGAGTGACCAACTCTATCAGGTCATCCAAAGCTGGGAACACTAC	573
Qy <td>541</td> <td>cgagagtga</td> <td>549</td>	541	cgagagtga	549
Db <td>574</td> <td>CGAGAGTGA</td> <td>582</td>	574	CGAGAGTGA	582
RESULT	6		
AF154114			
LOCUS	AF154114	630 bp	mRNA
			01-JUN-2000

```

DEFINITION Rattus norvegicus peptidoglycan recognition protein PGRP (Pgrp)
ACCESSION AF154114
VERSION AF154114.1 GI:8132325
KEYWORDS Norway rat.
SOURCE Rattus norvegicus
ORGANISM Rattus norvegicus
REFERENCE 1 (bases 1 to 630)
AUTHORS Rehman,A., Teodecki,E.E. and Krueger,J.M.
TITLE Direct Submission
JOURNAL Submitted (25-MAY-1999) Veterinary Comparative Anatomy,
Pharmacology, and Physiology, Washington State University, P.O. Box
646520, Pullman, WA 99164, USA
FEATURES
    source
        1..630
            /organism="Rattus norvegicus"
            /strain="Sprague-Dawley"
            /db_xref="taxon:10116"
            /tissue_type="spleen"
    gene
        1..630
            /gene="Pgrp"
    CDS
        20..571
            /note="similar to Homo sapiens and Mus musculus
            peptidoglycan recognition protein; similar to Trichoplusia
            ni and Mus musculus TAG7"
            /codon_start=1
            /product="peptidoglycan recognition protein PGRP"
            /protein_id="AAF73252.1"
            /db_xref="GI:8132326"
            /translation="MLFAWAPFPALLGLADSCFVPSSEKALPSECKLKPVRV
            VVISHTAGSCSDSECEQARNVQLQKMGWCDVAYNPLIGEDGVYRGRWTIK
            GDHTGLWNPMSIGITFMGYSYHRVPAPRALRALNLKCGVSEGLFNSYEVKGRHD
            VQSTLSPGDQYLIQIWDHYRE"
BASE COUNT 134 a 193 c 166 g 137 t
ORIGIN

Query Match 82.5%; Score 452.8; DB 10; Length 630;
Best Local Similarity 90.0%; Pred. No. 8.6e-107;
Matches 497; Conservative 0; Mismatches 52; Indels 3; Gaps 1;

Qy 1 atgttgttgcctgtctc---tccttcctcctcctgtctgtctggaacacctcctgagtttc 57
Db 20 ATGTTGTTGGCTGGGCTCCCTTCCTCCCTCCCTGGGCTGGCAGACTCCGTGCTTTTC 79

Qy 58 atcgtgccccagtgagtgaggggccctgcacatccgagtgctctagcgcctcgtgggcac 117
Db 80 GTCGTGCCCGCAGTGAGTGGAAGSCCTGCCATCCGAGTGTCTCCAAGGCGCTGAAGAA 139

Qy 118 ccagttcgactggtgtgatctcacacacagccggcagctcttgcaacagccggactcc 177
Db 140 CCAGTCCCTACGTGGTGTATCTCACACACAGCCGCGACGCTTCTGAGGAGCCAGACTCC 199

Qy 178 tgtgaacagcagggcccgcaatgtgcagcataccacaagaatgagctggtcgtggtcgcat 237
Db 200 TGTGAACAGCAGGCCCGCAATGTGCAGCTTACCAAAATGAACACACTGGGCTGGTGCAT 259

Qy 238 gtacctacaacttccttattggagagacgggtcattgtctatgaaggccggagctggaac 297
Db 260 GTAGCCTACAACCTTCTCATTTGGAGAGATGGTATGTCTACGAAGCGCGAGGCTGGACC 319

Qy 298 atcaagggtgaccacacagggcccatctggaatcccatctgtctatttggcataccttcag 357
Db 320 ATCAAGGTTGACACACAGGGCCCATCTCGAACCCCATGTCTATCGGCATCACCTTCATG 379

Qy 358 ggggaacttcattgacgggtaccgcaagcggccctccgtgctgacctaaattcttg 417
Db 380 GGTGACTACTACACCGGGTACCTGCAAGGGGCTCTCCGGTGTGCCCTAAATCTTCG 439

```

```

Qy 418 gaatgtgggtgtctcggggcttctcctgagatccaactatgaagtcaaaagcacacgggat 477
Db 440 AAATGTGGGTGTCTGAGGGCTTCTGAGATCTAACTATGAGTCAAAAGGACATCGAGAT 499

Qy 478 gtgcaaaagcactctctcctcaggtgacaaactctatcaggtcatccaaagctgggaacac 537
Db 500 GTGCAAAAGCACTCTCTCTCCAGGTGACCAGCTCTCAGAGATCATCCAAAGCTGGGACCAC 559

Qy 538 taccgagagtga 549
Db 560 TACCGAGAGTGA 571

RESULT 7
LOCUS AF076483 690 bp mRNA PRI 15-AUG-1998
DEFINITION Homo sapiens peptidoglycan recognition protein precursor (PGRP)
ACCESSION AF076483
VERSION AF076483.1 GI:3342532
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 690)
AUTHORS Kang,D., Liu,G., Lundstrom,A., Gelius,E. and Steiner,H.
TITLE A peptidoglycan recognition protein in innate immunity conserved
from insects to humans
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 95 (17), 10078-10082 (1998)
MEDLINE 98374308
REFERENCE 2 (bases 1 to 690)
AUTHORS Kang,D., Liu,G., Lundstrom,A., Gelius,E. and Steiner,H.
TITLE Direct Submission
JOURNAL Submitted (08-JUL-1998) Microbiology, Stockholm University,
Stockholm S-106 91, Sweden
FEATURES
    source
        1..690
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /tissue_type="bone marrow"
    gene
        1..690
            /gene="PGRP"
    CDS
        45..635
            /gene="PGRP"
            /codon_start=1
            /product="peptidoglycan recognition protein precursor"
            /protein_id="AAC31822.1"
            /db_xref="GI:3342533"
            /translation="MSRRSMLLAWALPSLLRLGAQETDPACCSPIVPRNEWKALAS
            ECAQHLSLPLRYVVVSHTAGSSCNTPASCCQOARNVOHYHMKTLGWCVDVGNFLIGED
            GLVYEGRWNTGAHSGHLWNPMSIGISFMGNVMDRVTTPQAIRAAQLGACGVAAGA
            LRSNYVLKGRHDVORTLSPGNQLYHLIONWPHYKSP"
            sig_peptide 45..104
            /gene="PGRP"
            mat_peptide 105..632
            /note="putative"
            /gene="PGRP"
            /product="peptidoglycan recognition protein"
BASE COUNT 129 a 250 c 188 g 123 t
ORIGIN

Query Match 57.7%; Score 316.6; DB 9; Length 690;
Best Local Similarity 77.2%; Pred. No. 1.5e-71;
Matches 385; Conservative 0; Mismatches 114; Indels 0; Gaps 0;

Qy 44 cctcctcgagttctatcgtgcccgcagtgagtgaggccctgccatccgagtgctcta 103
Db 127 CTGTGTGACCCCATAGTGTGCCCGAGAGTGGAGGCCCTGGCATCAGATGCGCCC 186

Qy 104 gcgcctggggcaccagcttcgctacgttggtgatctcacacacagccggcagcttctga 163

```

Qy	104	gcgcgtggtggccaccagttctgcagttggtgatctccacacagccggcagctctctgca	163
Db	172	AGCACCTGAGCCCTGCCCTTACGCTTATGTGGTGGTATCGCACACGGCGGCAGCAGCTGCA	231
Qy	164	acagcccgagactcctgtgaacagcagccgcgaatgtcagcattaccacaagaatgac	223
Db	232	ACACCCCGCCTCGTGCCACAGCAGCGCCGGAAATGTGCAGCACTACCACATGAAGACAC	291
Qy	224	tggctgtgtgcgtgtagcctcaaaccttctctattgtggagaggcgtgatgtctatgaag	293
Db	292	TGGCTGTGTCGACGTGGGCTACAACTTCTCTGATTGTGAGAAGACGGCTCGTATAGCAGG	351
Qy	284	gcgcaggtggaacatacaagggtgacacacacagggcccatctctgaatcccatgtctattg	343
Db	352	GCCGTGGCTGGAACTTCACGGGTGCCCACTTCAGGTCACTTATGGAACCCCATGCTCCATTG	411
Qy	344	gcataccttcattgggaactcatgagaccgggtaccgcgaagcggccctcgtgctg	403
Db	412	GCATCAGCTTCATGGGCACTCATGTGATCGGGTCCCAACCCAGGCCATCCGGGCAG	471
Qy	404	ccctaaactctctggaatgtggggtgtctcggggcttctctcaggtgacaaactatgaagtc	463
Db	472	CCCAGGCTCTACTGGCGTGGGTGTGGCTCAGGAGGCCCTGAGGTCCAACATATGTGCTCA	531
Qy	464	aagacacccggatgtgcaagaacactctctcctcaggtgacaaactctcatcaggtcatcc	523
Db	532	AAGGACACCCGGATGTGCAGCGTACACTCTCTCAGGCAACCACTCTACCACTCATCC	591
Qy	524	aaagctgggaacactaccg	542
Db	592	AGAATTGGCCACACTACCG	610

RESULT 9

CDR131676

LOCUS CDR131676 700 bp mRNA MAM 21-DEC-2000

DEFINITION Camelus dromedarius mRNA for peptidoglycan recognition protein.

ACCESSION AJ131676

VERSION AJ131676.1 GI:11990123

KEYWORDS peptidoglycan recognition protein.

SOURCE Arabian camel.

ORGANISM Camelus dromedarius

REFERENCE 1. (bases 1 to 700)

AUTHORS Kappeler,S.R., Farah,Z. and Puhan,Z.

TITLE Milk as a Source of Camel (Camelus dromedarius) Peptidoglycan Recognition Protein

JOURNAL Unpublished

REFERENCE 2. (bases 1 to 700)

AUTHORS Kappeler,S.R.

TITLE Direct Submission

JOURNAL Submitted (21-DEC-1998) Kappeler S.R., Institute of Food Science, Swiss Federal Institute of Technology, LFO F26, 8092 Zurich, SWITZERLAND

FEATURES

source

1. .700

/organism="Camelus dromedarius"

/sub_species="Somali breed"

/db_xref="taxon:9838"

/tissue_type="lactating mammary gland"

25. .606

/codon_start=1

/product="peptidoglycan recognition protein"

/protein_id="CAC19553.1"

/db_xref="GI:11990124"

/translation="MTRHCVLLVWALLLSLGAAREDPACGSIVPRRWRALASEC
RERLTPRVVYVSHATGSHCDPFPASCAQQAQNVQSHYRNLGWCVDGYNFLIGEDGL
VYEGRWNIKGAHGTWNPISIGISFMGNMVRPPPRALRAAQNLLACGVALGALR
SNVEYGRHDRVQPTLSFGDRLYEIIQTWSHYRA"

25. .87

sig_peptide

mat_peptide

88. .603

/product="peptidoglycan recognition protein"

007000

Pacleib, J., Paragas, V., Park, S., Patel, S., Pfeiffer, B.,
 Phouanavong, S., Pittman, G.S., Puri, V., Richards, S., Scheeler, F.,
 Stapleton, M., Strong, R., Swirskas, R., Tector, C., Williams, S.M.,
 Zaveri, J.S., Smith, H.O., Rubin, G.M. and Venter, J.C.
 Sequencing of Drosophila chromosome 2R, region 44D2-44E1
 Unpublished
 2 (bases 1 to 166867)
 Celniker, S.E., Agbayani, A., Arcaina, T.T., Baxter, E., Blazej, R.G.,
 Butenhorff, C., Champe, M., Chavez, C., Chew, M., Ciesiolka, L.,
 Doyle, C.M., Farfan, D.E., Galle, R., George, R.A., Harris, N.L.,
 Hoskins, R.A., Houston, K.A., Hummasti, S.R., Karra, K., Kearney, L.,
 Kim, E., Lee, B., Lewis, S., Li, P., Lomotan, M.A., Mazda, P.,
 Moshrefi, A.R., Moshrefi, M., Nixon, K., Pacleib, J.M., Park, S.,
 Pfeiffer, B., Poon, L., Sequeira, A., Sethi, H., Snir, E.,
 Swirskas, R.R., Wan, K.H., Weinburg, T., Zhang, R., Zieran, L.L. and
 Rubin, G.M.
 Direct Submission
 Submitted (16-MAR-1999) Drosophila Genome Center, Lawrence Berkeley
 Laboratory, MS 64-121, Berkeley, CA 94720, USA
 On Mar 17, 2001 this sequence version replaced gi:5670628.
 Sequence submitted by:
 Berkeley Drosophila Genome Project
 Lawrence Berkeley National Laboratory, MS 64-121
 Berkeley, CA 94720
 This sequence was assembled using end sequences from a whole genome
 shotgun and from subclones of this BAC and its neighboring clones.
 For further information about this sequence, including its location
 and relationship to other sequences, please visit our sequence
 archive Web site (<http://www.fruitfly.org/sequence/>) or send email
 to bdg@fruitfly.berkeley.edu.
 Location/Qualifiers
 1. 166867
 /organism="Drosophila melanogaster"
 /strain="y; cn bw sp"
 /db_xref="taxon:7227"
 /chromosome="2R"
 /map="44D2-44E1"
 /clone_lib="RPGI-98 (Roswell Park Cancer Institute
 Drosophila melanogaster BAC library, partial EcoRI in
 pRACE3.6)"
 BASE COUNT 46731 a 36967 c 36753 g 46416 t
 ORIGIN
 Query Match 24.0%; Score 131.6; DB 3; Length 166867;
 Best Local Similarity 56.0%; Pred. No. 5e-24;
 Matches 270; Conservative 0; Mismatches 209; Indels 3; Gaps 1;
 QY 57 catgtgccccgcagtgagtgagggccctgccatccagtgctctagccgcctggggca 116
 Db 159280 CATCATCTCCAAAGTCGGAGTGGGGCGCGGTTCGCCACGACGACGACCTCGCTGGCCAA 159339
 QY 117 cccagtcgcagtggtgatctcacacacagcggcgagcttctgcaacagccggactc 176
 Db 159340 CTACTCTGAGTCACCGCTGATCCACACACGCGTGGAAACTACTGACGACCAAGGCGCG 159399
 QY 177 ctgtgaacagcagccgcgaatgtgcagcattaccacaagaatgagctgggtggtgcga 236
 Db 159400 CTGCATCACAGCTGAGACATCCAGGCTCTACCATGACTCCCTGGGCTGGCCGA 159459
 QY 237 tgtagctacaactctctatttgagaggacggttcattgtctatgaagccgagctggaa 296
 Db 159460 TATCGGCTACAACTTCCTGATCGCGGGAGAGCGCAACGCTGTACGAGGTCGCGCTGGAA 159519
 QY 297 cataaagggtagcacacacagggcccatctggaatcccatgctctattgcatcaacttcat 356
 Db 159520 CGTTATGGGTGCTCAC---GGCACTTAACCTGGAACCTCAAGTCTATCGGCATCTCTTCT 159576
 QY 357 ggggaactcttgaccgggggtaccgcgcaagcgggccctcgctgctgcctaaatctctt 416
 Db 159577 GGGCAACTACAAATACCAACACCTCACCTCTGCTCAGATCACCCTGCCAGGCTCTGCT 159636

QY 417 ggaatgtgggtgtctcggggcttctcgtgagatcccaactatgaagtcacacagcaccggga 476
 Db 159637 CTCCTGATGGGTGATGTCGGGGCCAGATGTCCTCCGGATACATCTGTACGGACATGGCA 159696
 QY 477 tgtgcaaaagcactctctctccaggtgaccacactctcaggttcattcaggtcgggaaca 536
 Db 159697 GGTGGCTCCACCGAGTGCCTCCGGGACCAACATCTGGAACGAGATCCGACCTGGTCCA 159756
 QY 537 ct 538
 Db 159757 CT 159758
 RESULT 12
 AC007303/c
 LOCUS
 DEFINITION
 AC007303
 AC007303.7 GI:13702792
 HTG.
 KEYWORDS
 SOURCE
 ORGANISM
 Drosophila melanogaster
 fruit fly.
 Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
 Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 1 (bases 1 to 169509)
 Celniker, S.E., Adams, M.D., Kronmiller, B., Tyler, D., Wan, K.H.,
 Holt, R.A., Evans, C.A., Gocayne, J.D., Amanatides, P.G., Brandon, R.C.,
 Rogers, Y., An, H., Baldwin, D., Banzon, J., Beeson, K.Y., Busam, D.A.,
 Carlson, J.W., Center, A., Champe, M., Davenport, L.B., Dietz, S.M.,
 Dodson, K., Dorsett, V., Doup, L.E., Doyle, C., Dresnek, D., Farfan, D.,
 Ferrier, S., Frise, E., Galle, R.F., Garg, N.S., George, R.A.,
 Gonzalez, M., Houck, J., Hoskins, R.A., Hostin, D., Howland, T.J.,
 Ibegwam, C., Jallali, M., Kruse, D., Li, P., Mattel, B., Moshrefi, A.,
 McIntosh, T.C., Moy, M., Murphy, B., Nelson, C., Nelson, K.A., Nunoo, J.,
 Pacleib, J., Paragas, V., Park, S., Patel, S., Pfeiffer, B.,
 Phouanavong, S., Pittman, G.S., Puri, V., Richards, S., Scheeler, F.,
 Stapleton, M., Strong, R., Swirskas, R., Tector, C., Williams, S.M.,
 Zaveri, J.S., Smith, H.O., Rubin, G.M. and Venter, J.C.
 Sequencing of Drosophila chromosome 2R, region 44E-44F
 Unpublished
 2 (bases 1 to 169509)
 Celniker, S.E., Agbayani, A., Arcaina, T.T., Baxter, E., Blazej, R.G.,
 Butenhorff, C., Champe, M., Chavez, C., Chew, M., Ciesiolka, L.,
 Doyle, C.M., Farfan, D.E., Galle, R., George, R.A., Harris, N.L.,
 Hoskins, R.A., Houston, K.A., Hummasti, S.R., Karra, K., Kearney, L.,
 Kim, E., Lee, B., Lewis, S., Li, P., Lomotan, M.A., Mazda, P.,
 Moshrefi, A.R., Moshrefi, M., Nixon, K., Pacleib, J.M., Park, S.,
 Pfeiffer, B., Poon, L., Sequeira, A., Sethi, H., Snir, E.,
 Swirskas, R.R., Wan, K.H., Weinburg, T., Zhang, R., Zieran, L.L. and
 Rubin, G.M.
 Direct Submission
 Submitted (14-APR-1999) Drosophila Genome Center, Lawrence Berkeley
 Laboratory, MS 64-121, Berkeley, CA 94720, USA
 On Apr 21, 2001 this sequence version replaced gi:6435866.
 Sequence submitted by:
 Berkeley Drosophila Genome Project
 Lawrence Berkeley National Laboratory, MS 64-121
 Berkeley, CA 94720
 This sequence was assembled using end sequences from a whole genome
 shotgun and from subclones of this BAC and its neighboring clones.
 For further information about this sequence, including its location
 and relationship to other sequences, please visit our sequence
 archive Web site (<http://www.fruitfly.org/sequence/>) or send email
 to bdg@fruitfly.berkeley.edu.
 Location/Qualifiers
 1. 169509
 /organism="Drosophila melanogaster"
 /strain="y; cn bw sp"
 /db_xref="taxon:7227"
 /chromosome="2R"
 /map="44E-44F"

FEATURES
source

```

/clones="BACR43104 (D581)"
/clone_lib="RPC1-98 (Roswell Park Cancer Institute
Drosophila melanogaster BAC library, partial EcoRI in
pBACe3.6)"
BASE COUNT 46104 a 38440 c 38294 g 46671 t
ORIGIN

Query Match 24.0%; Score 131.6; DB 3; Length 169509;
Best Local Similarity 56.0%; Pred. No. 5e-24;
Matches 270; Conservative 0; Mismatches 209; Indels 3; Gaps 1;

Qy 57 catcgtgcccgcagtgagtgagggccctgccatccagtgctcttagcgcctggggca 116
||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 105920 CATCATCTCCAAAGTCGGAGTGGGGCGCGCTTCCGCCACGAGCAAGACTCGCTGCCAA 105861

Qy 117 cccagttcgtctacgttggtgatctcacacacagcggcgagcttctgcaacagccggga 176
||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 105860 CTACTGAGCTACGCCGCTGATCCACCAACCGCTGGAAACTACTGCAGCAGGCGCGC 105801

Qy 177 ctgtgaacagcagggcccgcaatgtgcagcattaccacaagaatgagctgggtgctgca 236
||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 105800 CTGCATCACAGCTGCAGACATCCAGGCTTACCACATGAGTCCCTTGGGCTGGCGCA 105741

Qy 237 ttagtcctacaactccttatttgagagggagcgtgctatgtatgaagcccgaggtggaa 296
||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 105740 TATCGGCTTACAACCTCTGATCGCGGAGAGCGGCAACGCTGTACGAGGCTCGCGCTGAA 105681

Qy 297 catcaagggtgacacacagggcccatctggaatcccatgtctattgcatcacccttcat 356
||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 105680 CGTATGAGGTGCTCAC---GCCACTTAAGTGAACCTCCAAAGTCTATFCGGCATCTCCTTCT 105624

Qy 357 ggggaactctcatggaccgggtaccgcgaagcgggccctcgctgctccctaaatcttct 416
||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 105623 GGGCACTTACATACCAACCCCTACCTCTGCTCAGATCACCCTGCGGCTGCAAGGCTCTCT 105564

Qy 417 ggaatgtgggtgtctcgtgggtctctgagatcccaactatgaagcacaagacacggga 476
||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 105563 CTCGGATCGGTGCTAGTCGCGGCGCAGATCGTTTCCGGATACATCTCTGTACGGACATCGGCA 105504

Qy 477 tgtgcaaaagcactctctctcagtgagtgacaaactctatcaggtcatcctcaaaagctggga 536
||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 105503 GGTGGCTCCACCGAGTCCCGGCGCACCAACATCTGGAACGAGATCCCGACCTGTGCCAA 105444

Qy 537 ct 538
||
Db 105443 CT 105442

RESULT 13
AC020451 172838 bp DNA HTG 03-JAN-2000
LOCUS Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
DEFINITION pieces.
ACCESSION AC020451
VERSION AC020451.1 GI:6664446
KEYWORDS HTG; HTGS_PHASE2.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 172838)
Direct Submission
Submitted (30-DEC-1999) Celera Genomics, 45 West Gude Drive,
Rockville, MD, USA
This sequence was identified as CDM:10213218 by the submitter.
For more information on this record e-mail to fly@celera.com.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

FEATURES             Location/Qualifiers
     source            1..172838
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
BASE COUNT 48031 a 38482 c 38368 g 47957 t
ORIGIN

Query Match 24.0%; Score 131.6; DB 2; Length 172838;
Best Local Similarity 56.0%; Pred. No. 5e-24;
Matches 270; Conservative 0; Mismatches 209; Indels 3; Gaps 1;

Qy 57 catcgtgcccgcagtgagtgagggccctgccatccagtgctcttagcgcctggggca 116
||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 150251 CATCATCTCCAAAGTCGGAGTGGGGCGCGCTTCCGCCACGAGCAAGACTCGCTGCCAA 150310

Qy 117 cccagttcgtctacgttggtgatctcacacacagcggcgagcttctgcaacagccggga 176
||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 150311 CTACTGAGCTACGCCGCTGATCCACCAACCGCTGGAAACTACTGCAGCAGGCGCGC 150370

Qy 177 ctgtgaacagcagggcccgcaatgtgcagcattaccacaagaatgagctgggtgctgca 236
||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 150371 CTGCATCACAGCTGCAGACATCCAGGCTTACCACATGAGTCCCTTGGGCTGGCGCA 150430

Qy 237 ttagtcctacaactccttatttgagagggagcgtgctatgtatgaagcccgaggtggaa 296
||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 150431 TATCGGCTTACAACCTCTGATCGCGGAGAGCGGCAACGCTGTACGAGGCTCGCGCTGAA 150490

Qy 297 catcaagggtgacacacagggcccatctggaatcccatgtctattgcatcacccttcat 356
||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 150491 CGTATGAGGTGCTCAC---GCCACTTAAGTGAACCTCCAAAGTCTATFCGGCATCTCCTTCT 150547

Qy 357 ggggaactctcatggaccgggtaccgcgaagcgggccctcgctgctccctaaatcttct 416
||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 150548 GGGCACTTACATACCAACCCCTACCTCTGCTCAGATCACCCTGCGGCTGCAAGGCTCTCT 150607

Qy 417 ggaatgtgggtgtctcgtgggtctctgagatcccaactatgaagcacaagacacggga 476
||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 150608 CTCGGATCGGTGCTAGTCGCGGCGCAGATCGTTTCCGGATACATCTCTGTACGGACATCGGCA 150667

Qy 477 tgtgcaaaagcactctctctcagtgagtgacaaactctatcaggtcatcctcaaaagctggga 536
||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 150668 GGTGGCTCCACCGAGTCCCGGCGCACCAACATCTGGAACGAGATCCCGACCTGTGCCAA 150727

Qy 537 ct 538
||
Db 150728 CT 150729

RESULT 14
AE003836 253273 bp DNA INV 04-OCT-2000
LOCUS Drosophila melanogaster genomic scaffold 14200001386047 section 7
DEFINITION of 52, complete sequence.
ACCESSION AE003836 AE002787
VERSION AE003836.2 GI:1072714
KEYWORDS HTG.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 253273)
Adams,M.D., Celniker,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D.,
Ananidis,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Galie,R.F.,
George,R.A., Lewis,S.E., Richards,S., Ashburner,M., Henderson,S.N.,
Sutton,G.G., Wortman,J.R., Vandell,M.D., Zhang,Q., Chen,L.X.,
Brandon,R.C., Rogers,Y.H., Blazej,R.G., Champe,M., Pfeiffer,B.D.,
Wan,K.H., Doyle,C., Baxter,E.G., Helt,G., Nelson,C.R., Gabor
Miklos,G.L., Abril,J.F., Agbayani,A., An,H.J.,
Andrews-Pfannkuch,C., Baldwin,D., Ballew,R.M., Beeson,K.Y.,
Baxendale,J., Bayraktaroglu,L., Beasley,E.M., Beeson,K.Y.,
Benos,P.V., Berman,B.P., Bhandari,D., Bolshakov,S., Borkova,D.,

```

Botchan, M.R., Bouck, J., Brokstein, P., Brottier, P., Brottier, P., Burtis, K.C., Busan, D.A., Butler, H., Cadieu, E., Center, A., Chandra, I., Cherry, J.M., Cawley, S., Dahike, C., Davenport, L.B., Davies, P., de Pablo, B., Delcher, A., Deng, Z., Mays, A.D., Dew, I., Dietz, S.M., Dodson, K., Doup, L.E., Downes, M., Dugan-Rocha, S., Dunkov, B.C., Dunn, P., Durbin, K.J., Evangelista, C.C., Ferraz, C., Ferriera, S., Fleischmann, W., Fosler, C., Gabriellian, A.E., Garg, N.S., Gelbart, W.M., Glasser, K., Glodek, A., Gong, F., Gorrell, J.H., Gu, Z., Guan, P., Harris, M., Harris, N.L., Harvey, D., Heiman, T.J., Hernandez, J.R., Houck, J., Hostin, D., Houston, K.A., Howland, T.J., Wei, M.H., thegam, C., Jalali, M., Kalush, F., Karpen, G.H., Ke, Z., Kennison, J.A., Ketchum, K.A., Kimmel, B.E., Kodira, C.D., Kraft, C., Kravitz, S., Kulp, D., Lai, Z., Lasko, P., Lei, Y., Levitsky, A.A., Li, J., Li, Z., Liang, Y., Lin, X., Liu, X., Mattei, B., McIntosh, T.C., McLeod, M.P., McPherson, D., Merkulov, G., Milshina, N.V., Mobarry, C., Morris, J., Moshrefi, A., Mount, S.M., Moy, M., Murphy, B., Murphy, L., Muzny, D.M., Nelson, D.L., Nelson, D.R., Nelson, K.A., Nixon, K., Nusskern, D.R., Pacleb, J.M., Palazzolo, M., Pittman, G.S., Pan, S., Pollard, J.D., Puri, V., Reese, M.G., Reinert, K., Remington, K., Saunders, J.R., Scheeler, F., Shen, H., Shue, B.C., Siden-Kiamos, I., Simpson, M., Skupski, M.P., Smith, T., Spier, E., Spradling, A.C., Stapleton, M., Strong, R., Sun, E., Svirska, R., Tector, C., Turner, R., Venter, E., Wang, A.H., Wang, X., Wang, Z.Y., Wassarman, D.A., Weinstock, G.M., Weissbach, J., Williams, S.M., Woodage, T., Worley, K.C., Wu, D., Yang, S., Yao, Q.A., Ye, J., Yeh, R.F., Zaveri, J.C., Zhu, M., Zhang, G., Zhao, Q., Zheng, L., Zheng, X.H., Zhong, F.N., Zhong, W., Zhou, X., Zhu, S., Zhu, X., Smith, H.O., Gibbs, R.A., Myers, E.W., Rubin, G.M. and Venter, J.C.

The genome sequence of *Drosophila melanogaster*
 Science 287 (5461), 2185-2195 (2000)
 20196006
 2 (bases 1 to 253273)
 Adams, M.D., Celniker, S.E., Gibbs, R.A., Rubin, G.M. and Venter, C.J.
 Direct Submission
 Submitted (21-04-2000) Celera Genomics, 45 West Gude Drive,
 Rockville, MD, USA
 On Oct 9, 2000 this sequence version replaced gi:7304001.
 Location/Qualifiers
 1. 253273
 /organism="Drosophila melanogaster"
 /db_xref="taxon:7227"
 /chromosome="2R"
 complement(<8774..>9076)
 /gene="CG12780"
 /product="CT36767"
 /db_xref="FLYBASE:FBan0012780"
 /db_xref="FLYBASE:FBgn0033301"
 complement(<8774..>9076)
 /gene="CG12780"
 /map="44C6-44C6"
 /db_xref="FLYBASE:FBan0012780"
 /db_xref="FLYBASE:FBgn0033301"
 complement(<8774..>9076)
 /gene="CG12780"
 /note="CG12780 gene product"
 /codon_start=1
 /db_xref="FLYBASE:FBan0012780"
 /db_xref="FLYBASE:FBgn0033301"
 /protein_id="AAF59079.1"
 /db_xref="GI:7304037"
 /translation="MSGYQVPLARVTSRRRGFEVSDDEPGISLFGFHGRNPEIVD
 LGNTWAADIGKDKDGRWYTNRDVELKGDVLIYWTYVYNGRDYHRNQSAGF"
 complement(21167..22519)
 /gene="Cyp6a14"
 /product="CT6361"
 /db_xref="FLYBASE:FBan0008687"
 /db_xref="FLYBASE:FBgn0033302"
 complement(<21167..>22519)
 /gene="Cyp6a14"
 /note="CG8687"
 /map="44C6-44D1"
 /db_xref="FLYBASE:FBan0008687"
 /db_xref="FLYBASE:FBgn0033302"

CDS
 complement(21176..22489)
 /gene="Cyp6a14"
 /note="Cyp6a14 gene product"
 /codon_start=1
 /db_xref="FLYBASE:FBan0008687"
 /db_xref="FLYBASE:FBgn0033302"
 /protein_id="AAF59078.1"
 /db_xref="GI:7304036"
 /translation="MIKDFSYFODRGAFNPRDDLTGHLFALGEGERAMRHKLTPV
 FTSGKIKQMSKVIVDGLRGDAMDKAKEKVEEGNVEIKDKLCARETTDVGSCAFG
 LECNSLDPSAEPRQKRGREIFTRRRSTVQSFIPTNARLARKLRKLVDPDDLTFQFM
 STVKNVDYRLKNGIKRNDIFIEOMIELRAEDQEAARKGQIGDLSHGLTLEQMAQAFV
 FVAGETSSSTMSLCYLALQPDIOQRREIESVLANDGGLVNDVLAOMTYLD
 QVLSGYSALAKAQDTFKPKPLSETLRKHPLPHLIRETTKDYQIPNSDIVLDKGILA
 LIPVHNHDDPEIYPEPFDPFDPPEVKNRHPRMAYLFFGDPKNCIGLRFKGIQA
 KIGLXLLRRRFSVSNRTDVLIFSKKSLTTNDGIYLYLKVVR"
 join(<26235..26464,26685..>26946)
 /gene="Cyp6a15Psi"
 /product="CT34554"
 /db_xref="FLYBASE:FBan0014753"
 /db_xref="FLYBASE:FBgn0033303"
 <26235..>26946
 /gene="Cyp6a15Psi"
 /note="CG14753"
 /db_xref="FLYBASE:FBan0014753"
 /db_xref="FLYBASE:FBgn0033303"
 join(26235..26464,26685..26946)
 /gene="Cyp6a15Psi"
 /note="Cyp6a15Psi gene product"
 /codon_start=1
 /db_xref="FLYBASE:FBan0014753"
 /db_xref="FLYBASE:FBgn0033303"
 /protein_id="AAF59077.1"
 /db_xref="GI:7304035"
 /translation="MKILRDDLDTFVSUVKPAISGMLTWTSPSPGGRSSRQSGSKPD
 LSHNWTLEQMAQAIYFFLAFETSSSTMSNCFRVTNELIVKGTSLILPVSFVHY
 DPHLYPHKLFDSRLKAYKSNRHPFAYLFFGTGPRSCIGLRFKGMQAKIGIVSAL
 QVR"
 27350..>28836
 /gene="Cyp6a13"
 /product="CT7990"
 /db_xref="FLYBASE:FBan0002397"
 /db_xref="FLYBASE:FBgn0033304"
 <27350..>28836
 /gene="Cyp6a13"
 /note="CG2397"
 /map="44C6-44D1"
 /db_xref="FLYBASE:FBan0002397"
 /db_xref="FLYBASE:FBgn0033304"
 27355..28836
 /gene="Cyp6a13"
 /note="Cyp6a13 gene product"
 /codon_start=1
 /db_xref="FLYBASE:FBan0002397"
 /db_xref="FLYBASE:FBgn0033304"
 /protein_id="AAF59076.1"
 /db_xref="GI:7304034"
 /translation="MLTLLVLVFTVGLLIYVYKLSWYSWSSRGVAGERPVPYFRGNMS
 GLGRLDHTDINLRIYKFRGVERVYCGYFTWTSLFIMDELIRDIRDESSPADR
 GLFNVRDDPLTGNLLFDGPEWRWLQNLTQVFTSGMKMFPMNVEVGEKLTQACR
 LQVGEIRAKDLCAFTTDVIGSCAFGLCNLSQDPESQFRMRGVSQPELHSLVQA
 FMAQPELRKLRFRLFRPEVSEFLDTVQTLDRRRENIHRNDLIQLLMELGEGV
 KDALSFQIAQALVFLAGDFTSTTMSFCLYELANPDQVRLRVLAVLRNNQ
 KLTVDSQVEMPYLDQVVAETLRKYPIPLHLLRSTKKEYQIPNSNLIPEGSKIIPVH
 SIHDDPELPDPKPFDFSEFEPEIKARHPAYLFFGDPKNCIGERGKQLQVKVGLV
 YLLHDDPFSRSEKTIQIPKFSRNFLRSTQGVHLRMEGLERP"
 complement(join(<30499..30823,30901..31084,31413..31618,
 31681..31628,31892..32209,32299..32640,33009..34211,
 34949..>35081))
 /gene="CG14748"
 /product="CT34549"
 /db_xref="FLYBASE:FBan0014748"
 /db_xref="FLYBASE:FBgn0033305"

gene
 complement(21176..22489)
 /gene="Cyp6a14"
 /note="Cyp6a14 gene product"
 /codon_start=1
 /db_xref="FLYBASE:FBan0008687"
 /db_xref="FLYBASE:FBgn0033302"
 /protein_id="AAF59078.1"
 /db_xref="GI:7304036"
 /translation="MIKDFSYFODRGAFNPRDDLTGHLFALGEGERAMRHKLTPV
 FTSGKIKQMSKVIVDGLRGDAMDKAKEKVEEGNVEIKDKLCARETTDVGSCAFG
 LECNSLDPSAEPRQKRGREIFTRRRSTVQSFIPTNARLARKLRKLVDPDDLTFQFM
 STVKNVDYRLKNGIKRNDIFIEOMIELRAEDQEAARKGQIGDLSHGLTLEQMAQAFV
 FVAGETSSSTMSLCYLALQPDIOQRREIESVLANDGGLVNDVLAOMTYLD
 QVLSGYSALAKAQDTFKPKPLSETLRKHPLPHLIRETTKDYQIPNSDIVLDKGILA
 LIPVHNHDDPEIYPEPFDPFDPPEVKNRHPRMAYLFFGDPKNCIGLRFKGIQA
 KIGLXLLRRRFSVSNRTDVLIFSKKSLTTNDGIYLYLKVVR"
 join(<26235..26464,26685..>26946)
 /gene="Cyp6a15Psi"
 /product="CT34554"
 /db_xref="FLYBASE:FBan0014753"
 /db_xref="FLYBASE:FBgn0033303"
 <26235..>26946
 /gene="Cyp6a15Psi"
 /note="CG14753"
 /db_xref="FLYBASE:FBan0014753"
 /db_xref="FLYBASE:FBgn0033303"
 join(26235..26464,26685..26946)
 /gene="Cyp6a15Psi"
 /note="Cyp6a15Psi gene product"
 /codon_start=1
 /db_xref="FLYBASE:FBan0014753"
 /db_xref="FLYBASE:FBgn0033303"
 /protein_id="AAF59077.1"
 /db_xref="GI:7304035"
 /translation="MKILRDDLDTFVSUVKPAISGMLTWTSPSPGGRSSRQSGSKPD
 LSHNWTLEQMAQAIYFFLAFETSSSTMSNCFRVTNELIVKGTSLILPVSFVHY
 DPHLYPHKLFDSRLKAYKSNRHPFAYLFFGTGPRSCIGLRFKGMQAKIGIVSAL
 QVR"
 27350..>28836
 /gene="Cyp6a13"
 /product="CT7990"
 /db_xref="FLYBASE:FBan0002397"
 /db_xref="FLYBASE:FBgn0033304"
 <27350..>28836
 /gene="Cyp6a13"
 /note="CG2397"
 /map="44C6-44D1"
 /db_xref="FLYBASE:FBan0002397"
 /db_xref="FLYBASE:FBgn0033304"
 27355..28836
 /gene="Cyp6a13"
 /note="Cyp6a13 gene product"
 /codon_start=1
 /db_xref="FLYBASE:FBan0002397"
 /db_xref="FLYBASE:FBgn0033304"
 /protein_id="AAF59076.1"
 /db_xref="GI:7304034"
 /translation="MLTLLVLVFTVGLLIYVYKLSWYSWSSRGVAGERPVPYFRGNMS
 GLGRLDHTDINLRIYKFRGVERVYCGYFTWTSLFIMDELIRDIRDESSPADR
 GLFNVRDDPLTGNLLFDGPEWRWLQNLTQVFTSGMKMFPMNVEVGEKLTQACR
 LQVGEIRAKDLCAFTTDVIGSCAFGLCNLSQDPESQFRMRGVSQPELHSLVQA
 FMAQPELRKLRFRLFRPEVSEFLDTVQTLDRRRENIHRNDLIQLLMELGEGV
 KDALSFQIAQALVFLAGDFTSTTMSFCLYELANPDQVRLRVLAVLRNNQ
 KLTVDSQVEMPYLDQVVAETLRKYPIPLHLLRSTKKEYQIPNSNLIPEGSKIIPVH
 SIHDDPELPDPKPFDFSEFEPEIKARHPAYLFFGDPKNCIGERGKQLQVKVGLV
 YLLHDDPFSRSEKTIQIPKFSRNFLRSTQGVHLRMEGLERP"
 complement(join(<30499..30823,30901..31084,31413..31618,
 31681..31628,31892..32209,32299..32640,33009..34211,
 34949..>35081))
 /gene="CG14748"
 /product="CT34549"
 /db_xref="FLYBASE:FBan0014748"
 /db_xref="FLYBASE:FBgn0033305"

gene
 complement(21176..22489)
 /gene="Cyp6a14"
 /note="Cyp6a14 gene product"
 /codon_start=1
 /db_xref="FLYBASE:FBan0008687"
 /db_xref="FLYBASE:FBgn0033302"
 /protein_id="AAF59078.1"
 /db_xref="GI:7304036"
 /translation="MIKDFSYFODRGAFNPRDDLTGHLFALGEGERAMRHKLTPV
 FTSGKIKQMSKVIVDGLRGDAMDKAKEKVEEGNVEIKDKLCARETTDVGSCAFG
 LECNSLDPSAEPRQKRGREIFTRRRSTVQSFIPTNARLARKLRKLVDPDDLTFQFM
 STVKNVDYRLKNGIKRNDIFIEOMIELRAEDQEAARKGQIGDLSHGLTLEQMAQAFV
 FVAGETSSSTMSLCYLALQPDIOQRREIESVLANDGGLVNDVLAOMTYLD
 QVLSGYSALAKAQDTFKPKPLSETLRKHPLPHLIRETTKDYQIPNSDIVLDKGILA
 LIPVHNHDDPEIYPEPFDPFDPPEVKNRHPRMAYLFFGDPKNCIGLRFKGIQA
 KIGLXLLRRRFSVSNRTDVLIFSKKSLTTNDGIYLYLKVVR"
 join(<26235..26464,26685..>26946)
 /gene="Cyp6a15Psi"
 /product="CT34554"
 /db_xref="FLYBASE:FBan0014753"
 /db_xref="FLYBASE:FBgn0033303"
 <26235..>26946
 /gene="Cyp6a15Psi"
 /note="CG14753"
 /db_xref="FLYBASE:FBan0014753"
 /db_xref="FLYBASE:FBgn0033303"
 join(26235..26464,26685..26946)
 /gene="Cyp6a15Psi"
 /note="Cyp6a15Psi gene product"
 /codon_start=1
 /db_xref="FLYBASE:FBan0014753"
 /db_xref="FLYBASE:FBgn0033303"
 /protein_id="AAF59077.1"
 /db_xref="GI:7304035"
 /translation="MKILRDDLDTFVSUVKPAISGMLTWTSPSPGGRSSRQSGSKPD
 LSHNWTLEQMAQAIYFFLAFETSSSTMSNCFRVTNELIVKGTSLILPVSFVHY
 DPHLYPHKLFDSRLKAYKSNRHPFAYLFFGTGPRSCIGLRFKGMQAKIGIVSAL
 QVR"
 27350..>28836
 /gene="Cyp6a13"
 /product="CT7990"
 /db_xref="FLYBASE:FBan0002397"
 /db_xref="FLYBASE:FBgn0033304"
 <27350..>28836
 /gene="Cyp6a13"
 /note="CG2397"
 /map="44C6-44D1"
 /db_xref="FLYBASE:FBan0002397"
 /db_xref="FLYBASE:FBgn0033304"
 27355..28836
 /gene="Cyp6a13"
 /note="Cyp6a13 gene product"
 /codon_start=1
 /db_xref="FLYBASE:FBan0002397"
 /db_xref="FLYBASE:FBgn0033304"
 /protein_id="AAF59076.1"
 /db_xref="GI:7304034"
 /translation="MLTLLVLVFTVGLLIYVYKLSWYSWSSRGVAGERPVPYFRGNMS
 GLGRLDHTDINLRIYKFRGVERVYCGYFTWTSLFIMDELIRDIRDESSPADR
 GLFNVRDDPLTGNLLFDGPEWRWLQNLTQVFTSGMKMFPMNVEVGEKLTQACR
 LQVGEIRAKDLCAFTTDVIGSCAFGLCNLSQDPESQFRMRGVSQPELHSLVQA
 FMAQPELRKLRFRLFRPEVSEFLDTVQTLDRRRENIHRNDLIQLLMELGEGV
 KDALSFQIAQALVFLAGDFTSTTMSFCLYELANPDQVRLRVLAVLRNNQ
 KLTVDSQVEMPYLDQVVAETLRKYPIPLHLLRSTKKEYQIPNSNLIPEGSKIIPVH
 SIHDDPELPDPKPFDFSEFEPEIKARHPAYLFFGDPKNCIGERGKQLQVKVGLV
 YLLHDDPFSRSEKTIQIPKFSRNFLRSTQGVHLRMEGLERP"
 complement(join(<30499..30823,30901..31084,31413..31618,
 31681..31628,31892..32209,32299..32640,33009..34211,
 34949..>35081))
 /gene="CG14748"
 /product="CT34549"
 /db_xref="FLYBASE:FBan0014748"
 /db_xref="FLYBASE:FBgn0033305"

gene
 complement(21176..22489)
 /gene="Cyp6a14"
 /note="Cyp6a14 gene product"
 /codon_start=1
 /db_xref="FLYBASE:FBan0008687"
 /db_xref="FLYBASE:FBgn0033302"
 /protein_id="AAF59078.1"
 /db_xref="GI:7304036"
 /translation="MIKDFSYFODRGAFNPRDDLTGHLFALGEGERAMRHKLTPV
 FTSGKIKQMSKVIVDGLRGDAMDKAKEKVEEGNVEIKDKLCARETTDVGSCAFG
 LECNSLDPSAEPRQKRGREIFTRRRSTVQSFIPTNARLARKLRKLVDPDDLTFQFM
 STVKNVDYRLKNGIKRNDIFIEOMIELRAEDQEAARKGQIGDLSHGLTLEQMAQAFV
 FVAGETSSSTMSLCYLALQPDIOQRREIESVLANDGGLVNDVLAOMTYLD
 QVLSGYSALAKAQDTFKPKPLSETLRKHPLPHLIRETTKDYQIPNSDIVLDKGILA
 LIPVHNHDDPEIYPEPFDPFDPPEVKNRHPRMAYLFFGDPKNCIGLRFKGIQA
 KIGLXLLRRRFSVSNRTDVLIFSKKSLTTNDGIYLYLKVVR"
 join(<26235..26464,26685..>26946)
 /gene="Cyp6a15Psi"
 /product="CT34554"
 /db_xref="FLYBASE:FBan0014753"
 /db_xref="FLYBASE:FBgn0033303"
 <26235..>26946
 /gene="Cyp6a15Psi"
 /note="CG14753"
 /db_xref="FLYBASE:FBan0014753"
 /db_xref="FLYBASE:FBgn0033303"
 join(26235..26464,26685..26946)
 /gene="Cyp6a15Psi"
 /note="Cyp6a15Psi gene product"
 /codon_start=1
 /db_xref="FLYBASE:FBan0014753"
 /db_xref="FLYBASE:FBgn0033303"
 /protein_id="AAF59077.1"
 /db_xref="GI:7304035"
 /translation="MKILRDDLDTFVSUVKPAISGMLTWTSPSPGGRSSRQSGSKPD
 LSHNWTLEQMAQAIYFFLAFETSSSTMSNCFRVTNELIVKGTSLILPVSFVHY
 DPHLYPHKLFDSRLKAYKSNRHPFAYLFFGTGPRSCIGLRFKGMQAKIGIVSAL
 QVR"
 27350..>28836
 /gene="Cyp6a13"
 /product="CT7990"
 /db_xref="FLYBASE:FBan0002397"
 /db_xref="FLYBASE:FBgn0033304"
 <27350..>28836
 /gene="Cyp6a13"
 /note="CG2397"
 /map="44C6-44D1"
 /db_xref="FLYBASE:FBan0002397"
 /db_xref="FLYBASE:FBgn0033304"
 27355..28836
 /gene="Cyp6a13"
 /note="Cyp6a13 gene product"
 /codon_start=1
 /db_xref="FLYBASE:FBan0002397"
 /db_xref="FLYBASE:FBgn0033304"
 /protein_id="AAF59076.1"
 /db_xref="GI:7304034"
 /translation="MLTLLVLVFTVGLLIYVYKLSWYSWSSRGVAGERPVPYFRGNMS
 GLGRLDHTDINLRIYKFRGVERVYCGYFTWTSLFIMDELIRDIRDESSPADR
 GLFNVRDDPLTGNLLFDGPEWRWLQNLTQVFTSGMKMFPMNVEVGEKLTQACR
 LQVGEIRAKDLCAFTTDVIGSCAFGLCNLSQDPESQFRMRGVSQPELHSLVQA
 FMAQPELRKLRFRLFRPEVSEFLDTVQTLDRRRENIHRNDLIQLLMELGEGV
 KDALSFQIAQALVFLAGDFTSTTMSFCLYELANPDQVRLRVLAVLRNNQ
 KLTVDSQVEMPYLDQVVAETLRKYPIPLHLLRSTKKEYQIPNSNLIPEGSKIIPVH
 SIHDDPELPDPKPFDFSEFEPEIKARHPAYLFFGDPKNCIGERGKQLQVKVGLV
 YLLHDDPFSRSEKTIQIPKFSRNFLRSTQGVHLRMEGLERP"
 complement(join(<30499..30823,30901..31084,31413..31618,
 31681..31628,31892..32209,32299..32640,33009..34211,
 34949..>35081))
 /gene="CG14748"
 /product="CT34549"
 /db_xref="FLYBASE:FBan0014748"
 /db_xref="FLYBASE:FBgn0033305"

```
/evidence=not_experimental
complement(<30499..>35081)
/gene="CG14748"
/map="44D1-44D1"
/db_xref="FLYBASE:FBan0014748"
/db_xref="FLYBASE:FBgn0033305"
/evidence=not_experimental
complement(join(30499..30823,30901..31084,31413..31618,
31681..31828,31892..32209,32299..32640,33009..34211,
34949..35081))
/gene="CG14748"
/note="CG14748 gene product"
/codon_start=1
/db_xref="FLYBASE:FBan0014748"
/db_xref="FLYBASE:FBgn0033305"
/evidence=not_experimental
/protein_id="AAF59075.1"
/db_xref="GI:7304033"
/translation="MGSYGLPLLLLRLLHGGKSFVPRHTDLSIIIVEDAK
ELFRVPLAQRIAPHFRRDNLRLHPDYVYERDRNAPSSRRSACODCSKGPVSSO
PYFYFGSGAGGVDRRDRDRDHYRPHSNTAVDHYRGPSPGSSSYERHPYSS
EPHGSSSSASSSTFFSGSGSSSGSYGFHTPPSPSYGGDHPYVDRHDSPHRYD
GPRDWTLPSSGSSAYRPPETDRIDRYPPYRPGSDYSPYRPPSPSHYLD
RDGAGPPGPNKPNKFPVYLVIGEDNWGHYGGSGSSSSSSSYWGLTESQ
RKQDQDFNYFOLGAPHRNSVLYPGVGGEDYPRRRNDYRQWTRAGDECSAK
TSEGLRHKTAVRHAYNVPLTECLSDRPSFVCHTFVYRQAGRDNCMLCDRP
INWLYDYEDRVDYDYSMADDMDVCRPPRRHDYAGPSTASIDPRNACFFRAI
DATRFKSLVRSVSGECMEICRSTKTRAFPRYQGRHAGVIDNCLSDW
PYRMDKRLHLIDKADFIRASVGHGCEIPIVDERKHKLCYLGYSPPARLLPPAI
KRVISVPYKACIKRFPETFPKCFASVYSGHSSFNCSLSDLDQTELKQDHYA
HTKDRDLFWLFWANNPDMTCRDKVNTIGSRVNRDRMDIFRPPGDVAMRHYTVSGKPC
```

Query Match 24.0%; Score 131.6; DB 3; Length 253273;
Best Local Similarity 56.0%; Pred. No. 4.7e-24;
Matches 270; Conservative 0; Mismatches 209; Indels 3; Gaps 1;

```
QY 57 catcgtgccgcagtgagtgaggccctgcctagcagtgctctagcgcctggggca 116
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 173856 CATCATCTCAAGTCGAGTGGGGCGCGCTTCGCCACGACGACCTCGCTGCCAA 173915

QY 117 cccagttcgctacgtggtgtctctcaacacagccgagcttcttgaacagcccgactc 176
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 173916 CTACCTGAGCTACGCGGTGATCCACCACCGCTGGAAACTACTGACGACCAAGCGCG 173975

QY 177 ctgtgaacacagccgcgaatgtgcaagattaccacaaagaatgagctgggtgcga 236
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 173976 CTGCATCACACAGCTGCAGAACATCCAGGCTACCATGACTCCCTGGGCTGGGCCGA 174035

QY 237 ttagcctcaacttccttattgagaggagcgtcatgtctatgaagccgagcgtgaa 296
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 174036 TATCGGCTACACTTCTGTATCGGCGGAGGACGCAAGCTGTACGAGGCTCGCGCTGAA 174095

QY 297 catcaaggtgacacacagggccactcttgaatcccatgtctattggcatcaccttcat 356
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 174096 CGTATGGTGTCTCAC---GCCACTAAGTGAAGTCAAGTCTATCGCATCTCTCTCTCT 174152

QY 357 ggggaacttcagcagcgggtaccgcaagcggccctcgctgctccctaaattctct 416
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 174153 GGGCAACTACAAATACCAACACCTTCCTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 174212

QY 417 ggaatgtgggggtgctcggggcttcctgagatcccaactgaagtcacaaagacacccgga 476
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 174213 CTCGATGCGGTGAGTCGCGGCCAGATCGTTCGGGATACATCTGTGTGTGTGTGTGTGT 174272

QY 477 tgtgcaagcactctctctcagtgaccacactctatcaggtctatccaaagcgtgggaaca 536
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 174273 GGTGCGGTCCACCGAGTGGCCCGCACCAACATCTGGAACGAGATCGGACCTGGTCTCAA 174332

QY 537 ct 538
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 174333 CT 174334
```

RESULT 15

AC007785/c

LOCUS

DEFINITION

AC007785

AC007785.1

GI:5042403

HTG.

KEYWORDS

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 166500)

Lamerdin,J.E., McCreedy,P.M., Richardson,P., Sakaldasis,G.,

Burkhardt-Schultz,K., Gordon,L., Scott,D., Johnson,G., Stillwagen,S.,

Phan,H., Velasco,N., Do,L., Regala,W., Terry,A., Danganan,L.,

Erlar,A., Christensen,M., Georgescu,A., Avila,J., Attix,C.,

Andreise,T., Amico-Keller,G., Coefield,J., Duarte,S., Lucas,S.,

Bruce,R., Thomas,P., Quan,G., Kronmiller,B., Arellano,A.,

Sanders,C., Ow,D., Nolan,M., Trong,S., Kobayashi,A., Olsen,A.S. and

Carrano,A.V.

Sequence analysis of a 2.5 Mb region in 19q13.2-13.3 between APOE

and CSRI

Unpublished

2 (bases 1 to 166500)

Lamerdin,J.E.

Direct Submission

Submitted (11-JUN-1999) Joint Genome Institute, Lawrence Livermore

National Laboratory, 7000 East Ave., Livermore, CA 94551, USA

Map and sequence oriented from q centromere to telomere. BAC 282485

overlaps cosmid r30477 to the left from bases 1 to 4,419 of this

accession, and overlaps BAC 264576 (CIT-B-297N14; AC006262) from

bases 146,203 to 166,500. Additional chromosome 19 map and sequence

information may be obtained at:

http://www-bio.lnl.gov/bbrp/genome/genome.html.

Location/Qualifiers

1. .166500

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="CIT-B-344H19"

/chromosome="19"

/map="19q13.3 between DM and c5R1"

/clone_lib="Cal Tech CIT-B BAC library"

/note="LLNL clone name: BC282485"

70. .145

/rpt_family="(TCAA)n"

complement(177..453)

/rpt_family="AluJb"

complement(472..712)

/rpt_family="MIR"

746..904

/note="DSS similarity to overlapping ESTs:

N39722 YX92604.r1 Homo sapiens cDNA clone 269214 5';

(12..170); 99% identity..AA283147 ztl4d07.r1 NCL_CGAP_GCB1

Homo sapiens cDNA clone IMAGE:713101 5'; (6..163); 99%

identity..AA143160 z149d07.r1 Soares pregnant uterus NbHPU

Homo sapiens cDNA clone 505261 5'; (1..43); 100%

identity..AA142881 z149d07.s1 Soares pregnant uterus NbHPU

Homo sapiens cDNA clone 505261 3'; (570..527); 64%

identity."

join(<746..904,1157..1239,4294..4451,5875..6036,

12784..12994,13080..13162,13903..13958,14053..14128,

14429..14568,14726..14804,14982..15039,15813..15885,

16025..16122)

/note="Hypothetical arginine-rich gene product;

Hypothetical CDS constructed from overlapping ESTs and

Xrail predictions. EST coverage is lacking for some

portions of the CDS. Gene may be alternatively spliced, as

multiple transcripts map into this region; alternatively,

these individual clones may represent artifactual events

during RT and cDNA creation"

/codon_start=3

FEATURES

source

repeat_region

repeat_region

repeat_region

misc_feature

CDS

Oy 224 tgggctgtgcgatgtagcctacaact 250
|||||
Db 20494 TGGGCTGGTGGCGACGTGGGCTACAAGT 20468

Search completed: December 17, 2001, 09:36:51
Job time: 6531 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 17, 2001, 08:42:11 ; Search time 1569.57 seconds
(without alignments)
5770.339 Million cell updates/sec

Title: US-09-462-625-1

Perfect score: 549

Sequence: 1 atgtttgttgctgtgtct.....gggaacataccgagagtga 549

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 1472140 seqs, 8248589755 residues

Word size : 0

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

GenEmbl:*

Result No.	Score	Query Match	Length	DB ID	Description
1	549	100.0	549	6	AR124884 Sequence
2	549	100.0	669	10	AF193843 Mus muscu
3	498	90.7	680	10	AF076482 Mus muscu
4	498	90.7	713	10	BC005582 Mus muscu
5	379	69.0	678	10	MMRNATMS1
6	248	45.2	4205	10	MMTAG7
7	40	7.3	630	4	AF154114 Rattus no
8	26	4.7	700	4	CDRI31676
9	22	4.0	690	9	AF076483 Homo sapi
10	22	4.0	724	9	AF242517 Homo sapi
11	22	4.0	166500	9	AC007785 Homo sapi
12	21	3.8	188095	4	AC091505 Sus scrofa
13	20	3.6	606	10	RNCST5
14	20	3.6	1789	10	AF346597 Rattus norv
15	20	3.6	1858	10	RNCST3
16	20	3.6	1927	10	RNCST1
17	20	3.6	1931	10	RSCALPST
18	20	3.6	1972	10	AF346598 Rattus no
19	20	3.6	2041	10	RNCST2
20	20	3.6	35877	9	AL392105 Human DNA
21	20	3.6	92711	2	AL390251 Homo sapi
22	20	3.6	106950	2	AC016474 Homo sapi
23	20	3.6	150472	2	AC021399 Homo sapi
24	20	3.6	178739	2	AC024688 Homo sapi
25	20	3.6	184452	2	AC023169 Homo sapi
26	20	3.6	187174	9	AC006241 Homo sapi
27	20	3.6	193889	2	AL513479 Homo sapi
28	20	3.6	202655	10	AC026682 Mus muscu
29	20	3.6	222016	2	AC023048 Homo sapi
30	19	3.5	1159	3	EHC6
31	19	3.5	3580	9	AK023747 Homo sapi
32	19	3.5	5809	9	AB007974 Homo sapi
33	19	3.5	6494	9	AB002323 Human mRNA
34	19	3.5	14279	10	RATDYNEINC
35	19	3.5	14351	10	AY004877 Mus muscu
36	19	3.5	15500	10	RATCDHC
37	19	3.5	31934	3	U64859 Caenorhabdi
38	19	3.5	118334	9	AC005075 Homo sapi
39	19	3.5	150356	2	AC068791 Homo sapi
40	19	3.5	153756	9	AC073866 Homo sapi
41	19	3.5	158634	2	AP000909 Homo sapi
42	19	3.5	165394	2	AP000877 Homo sapi
43	19	3.5	176805	2	AL391729 Homo sapi
44	19	3.5	179087	2	AC069489 Homo sapi
45	19	3.5	189135	2	AL596265 Mus muscu

ALIGNMENTS

RESULT 1

AR124884
LOCUS AR124884 549 bp DNA
DEFINITION Sequence 1 from patent US 6172211.
ACCESSION AR124884
VERSION AR124884.1 GI:14110245
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.

PAT 16-MAY-2001

REFERENCE 1 (bases 1 to 549)
AUTHORS Georgiev,G.P., Kiselev,S.L., Prokhorchouk,E.B. and Ostermann,E.
TITLE Nucleic acid encoding tag7 polypeptide
JOURNAL Patent: US 6172211-A 1 09-JAN-2001;
FEATURES Location/Qualifiers
source 1..549
/organism="unknown"

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

BASE COUNT 117 a 161 c 151 g 120 t
ORIGIN

SUMMARIES

Query Match 100.0%; Score 549; DB 6; Length 549;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 549; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atgttggctgtgcttcttcccttcccttgggttggcgaacctctgcagtttcatc 60
|||||
DB 1 ATGTGTTTGGCTGTCCTCTCCCTCCCTGGGTGCGCAACCTCTCGCAGTTTCATC 60
|||||
QY 61 gtgcccagcagtgagtgagggccctgcatccgagtgctctagcgcctggggcaccaca 120
|||||
DB 61 GTGCCCGCAGTGAGTGAGGGCCCTGCCATCCGAGTGTCTAGCCGCTGGGCACCCA 120
|||||
QY 121 gttcgtactgtgtgatctcacacagccggcagcttctgcaacgcccgaacctctgt 180
|||||
DB 121 GTTCGCTACGTGGTGTATCTCACACACAGCCGCGAGCTTCTGCAACAGCCGAGCTCCTGT 180
|||||
QY 181 gaacagcagccgcgaatgtgagcattaccacaagaatgagctggctggcgatgta 240
|||||
DB 181 GAACAGCAGCCGCAATGTGAGCAATATACCAAGAATGAGCTGGGCTGGTGGCATGTA 240
|||||
QY 241 gctcaaaccttcttattgagagacggtcatgtctatgaagggcagggcctggaacatc 300
|||||
DB 241 GCCTACACTTCTTATTGAGAGGAGCGTCTATGTAAGGCGGAGGCTGGAACATC 300
|||||
QY 301 aaggggtgaccacacagggcccatctggaatcccatgtctattggcacccttcattgggg 360
|||||
DB 301 AAGGGTGACACACAGGGCCCTCTGGAATCCCATGTCTATTGGCATCACTTCATGGGG 360
|||||
QY 361 aacttcagcaggggtacccgcaagggccctccgtcgtccctaaatctctggaa 420
|||||
DB 361 AACTTCATGGAGCGGTACCCGCAAGGGCCCTCCGTGCTGCCCTAAATCTCTGGAA 420
|||||
QY 421 tgtgggggtctctggggcttctgagatccaactatgaagtcaaaagacacccggatgtg 480
|||||
DB 421 TGTGGGGTCTCTCGGGCTCTCTGAGATCAACTATGAAGTCAAGGACACCGGGATGTG 480
|||||
QY 481 caaagcactctctccaggtgaccactctatcaggttcacaaagctgggaaactac 540
|||||
DB 481 CAAAGCACTCTCTCTCCAGGTGACCAACTCTATCAGTCTATCCAAAGCTGGGAACACTAC 540
|||||
QY 541 cgagagtga 549
|||||
DB 541 CGAGAGTGA 549
|||||

RESULT 2
AF193843 AF193843 669 bp mRNA ROD 07-NOV-1999
LOCUS Mus musculus TAG7-like protein mRNA, complete cds.
DEFINITION AF193843
ACCESSION AF193843
VERSION AF193843.1 GI:6273360
KEYWORDS house mouse.
SOURCE Mus musculus
ORGANISM Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 669)
AUTHORS Slayton,W.B., Rigaa,A., Hancock,J.D., Zaugg,J.K., Le,T.V.,
Trautman,M.S., Spangrude,G.J., Carroll,W.L. and Schibler,K.R.
TITLE Granulocyte-colony stimulating factor up-regulates expression of
murine tag7 during myeloid differentiation
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 669)
AUTHORS Slayton,W.B., Rigaa,A., Hancock,J.D., Zaugg,J.K., Le,T.V.,
Trautman,M.S., Spangrude,G.J., Carroll,W.L. and Schibler,K.R.
TITLE Direct Submission
JOURNAL Submitted (11-OCT-1999) Pediatrics; University of Utah, 50 North
Medical Drive, Room 2A126, Salt Lake City, UT 84132, USA
FEATURES
SOURCE 1. .669
/organism="Mus musculus"
/db_xref="taxon:10090"

Query Match 100.0%; Score 549; DB 10; Length 669;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 549; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atgttggctgtgcttcttcccttcccttgggttggcgaacctctgcagtttcatc 60
|||||
DB 25 ATGTGTTTGGCTGTCCTCTCCCTCCCTGGGTGCGCAACCTCTCGCAGTTTCATC 84
|||||
QY 61 gtgcccagcagtgagtgagggccctgcatccgagtgctctagcgcctggggcaccaca 120
|||||
DB 85 GTGCCCGCAGTGAGTGAGGGCCCTGCCATCCGAGTGTCTAGCCGCTGGGCACCCA 144
|||||
QY 121 gttcgtactgtgtgatctcacacagccggcagcttctgcaacgcccgaacctctgt 180
|||||
DB 145 GTTCGCTACGTGGTGTATCTCACACAGCCGCGAGCTCTGCAACAGCCGCGAGCTCCTGT 204
|||||
QY 181 gaacagcagccgcgaatgtgagcattaccacaagaatgagctgggtggcgatgta 240
|||||
DB 205 GAACAGCAGCCGCAATGTGAGCAATATGAGCAGTATACCAAGAATGAGCTGGGCTGGTGGCATGTA 264
|||||
QY 241 gctcaaaccttcttattgagagggcaggtcatgtctatgaagggcagggcctggaacatc 300
|||||
DB 265 GCCTACACTTCTTATTGAGAGGAGCGGTCTATGTAAGGCGGAGGCTGGAACATC 324
|||||
QY 301 aaggggtgaccacacagggcccatctggaatcccatgtctattggcacccttcattgggg 360
|||||
DB 325 AAGGGTGACACACAGGGCCCTCTGGAATCCCATGTCTATTGGCATCACTTCATGGGG 384
|||||
QY 361 aacttcagcaggggtacccgcaagggccctccgtcgtccctaaatctctggaa 420
|||||
DB 385 AACTTCATGGAGCGGTACCCGCAAGGGCCCTCCGTGCTGCCCTAAATCTCTGGAA 444
|||||
QY 421 tgtgggggtctctggggcttctgagatccaactatgaagtcaaaagacacccggatgtg 480
|||||
DB 445 TGTGGGGTGTCTCGGGCTCTCTGAGATCCAACTATGAAGTCAAAAGGACACCGGGATGTG 504
|||||
QY 481 caaagcactctctccaggtgaccactctatcaggttcacaaagctgggaaactac 540
|||||
DB 505 CAAAGCACTCTCTCTCCAGGTGACCAACTCTATCAGGTCTATCCAAAGCTGGGAACACTAC 564
|||||
QY 541 cgagagtga 549
|||||
DB 565 CGAGAGTGA 573
|||||

RESULT 3
AF076482 AF076482 680 bp mRNA ROD 15-AUG-1998
LOCUS Mus musculus peptidoglycan recognition protein precursor (Pgrp)
DEFINITION AF076482
ACCESSION AF076482
VERSION AF076482.1 GI:3342530
KEYWORDS house mouse.
SOURCE Mus musculus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 680)

AUTHORS	Kang,D., Liu,G., Lundstrom,A., Gelius,E. and Steiner,H.
TITLE	A peptidoglycan recognition protein in innate immunity conserved from insects to humans
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 95 (17), 10078-10082 (1998)
MEDLINE	98374308
REFERENCE	2 (bases 1 to 680)
AUTHORS	Kang,D., Liu,G., Lundstrom,A., Gelius,E. and Steiner,H.
TITLE	Direct Submission
JOURNAL	Submitted (08-JUL-1998) Microbiology, Stockholm University, Stockholm S-106 91, Sweden
FEATURES	Location/Qualifiers
source	1..680
gene	/organism="Mus musculus" /db_xref="taxon:10090" /tissue_type="spleen"
CDS	1..680 /gene="Pgrp" 36..584 /gene="Pgrp" /note="PGRP" /codon_start=1
sig_peptide	/product="peptidoglycan recognition protein precursor" /protein_id="AAC31821.1" /db_xref="GI:3342511"
mat_peptide	/translation="MLFACALLALLGLATSCSFLVPRSEWRALPSCSRLGHPVRYV VISHTPGSCFNPSDSCQARNVQVHKNELGMCDAINFLIGEDGVTYEGRWGNNIG DHTGTPWNPMSIGITTFMNFMDRVPAPKRALRAALNLECGVSRGLRSNYEVRGHRD QSLSPGDQLYQVIOISWEHYRE" 36..83
BASE COUNT	146 a 216 c 174 g 144 t
ORIGIN	
Query Match	90.7% ; Score 498 ; DB 10 ; Length 680 ;
Best Local Similarity	99.8% ; Pred. No. 2.7e-279 ;
Matches 548 ; Conservative	0 ; Mismatches 1 ; Indels 0 ; Gaps 0 ;
Qy 1	atgtgtttgctgtgctctcttgcctctcctgggtctggcaactcctcgcagttcctc 60
Db 36	ATGTTGTTGGCTGTGCTCTCTTGGCCCTCTGGGGTCTGGCAACCTCTCGAGCTTTCATC 95
Qy 61	gtgccccgcagtgagtgagggccctgccatcccgagtgctctagccgctggggcaccaca 120
Db 96	GTGCCCCGCAGTGAGTGGAGGGCCCTGCCATCCGAGTGTCTAGCGGCTGGGCATCCCA 155
Qy 121	gttcgctacgttggtgatctcacacagccggcagcttctgtcacagcccggaactcctgt 180
Db 156	GTTCCGCTACGTTGGTGATCTCACACAGCCGGCAGCTTCTGCAACAGCCCGGACTCCCTGT 215
Qy 181	gaacagcagggcccgcaatgtgcagcattaccacaagaatgagctgggctggtgcagtgta 240
Db 216	GAACAGAGGCCCCGCAATGTCCAGCATTTACCACAGAATGAGCTGGGCTGTGGCATGTA 275
Qy 241	gcctacaacttccttattggagagcaggtcatgtctatgaagccgagggctggaaacatc 300
Db 276	GCCTACAACTTCCTTATTGGAGAGACGGTTCATGTCTATGAAGCCGAGGCTGGAACATC 335
Qy 301	aagggtagcaacacagggcccatctggaatcccatgtctattggcatcaccttcattgggg 360
Db 336	AAGGGTAGCACACAGAGGCCCATCTGGAAATCCCATGTCTATTGGCATCACCTTCATGGGG 395
Qy 361	aacttcattgaccgggtaccgcgaagcgggccctccgtgctgcctcaaatcttcttgaa 420
Db 396	AACATTCATGGACGGGTACCCGCAAGGGGCCCTCCGTGCTGCCCPAAATCTTCTGAA 455
Qy 421	tgtggggtgtctcgggggttccttgagatcccaactatgaagtcaaggacacccggatgtg 480
Db 456	TGTGGGGTGTCTCGGGGCTTCTGAGATCCAACTATGAAGTCAAGGACACCGGGATGTG 515


```

Biologiy, Vavilova 34/5, 117334 Moscow, RUSSIA
2 (bases 1 to 4205)
AUTHORS Kiselev, S.L., Kustikova, O.S., Korobko, E.V., Prokhortchouk, E.B.,
Kabishv, A.A., Lukanidin, E.M. and Georgiev, G.P.
TITLE Molecular cloning and characterization of the mouse tag7 gene
encoding a novel cytokine
J. Biol. Chem. 273 (29), 18633-18639 (1998)
MEDLINE 98325081
COMMENT Related entry: X86374.
FEATURES
    source          Location/Qualifiers
                    1..4205
                        /organism="Mus musculus"
                        /strain="B6/CBAF1J"
                        /db_xref="taxon:10090"
                        /map="A3 cytogenetic band"
    gene            3507..3818
                        /gene="tag7"
    TATA_signal     3507..3513
                        /gene="tag7"
    mRNA            3533..3818
                        /gene="tag7"
    exon            3533..3818
                        /gene="tag7"
                        /number=1
    CDS              3571..3818
                        /gene="tag7"
                        /codon_start=1
                        /product="TAG7 protein"
                        /protein_id="CAA72803.1"
                        /db_xref="GI:2765148"
                        /db_xref="SPREMBL:O62185"
                        /translation="MLFACALLALGLATSCSFIVPRSEWRALPSECSSRLGHVRYV
                        VISHTAGSCNSPDSCSQARNVQHYHKNELGWCDVAY"
BASE COUNT 1013 a 1157 c 1020 g 1003 t 12 others
ORIGIN

```

Query Match	45.2%	Score 248	DB 10	Length 4205	
Best Local Similarity	100.0%	Pred. No. 3e-133			
Matches 248	Conservative 0	Mismatches 0	Indels 0	Gaps 0	

QY	1	atgtgtttccctgctctctcttgcctcctcggtgtgtgcaacctctcgagtttcac	60
Db			
QY	3571	ATGTTGTTTGCCTGTGCTTCTCTTGCCCTCTCGGCTCTGCCAACCCTCTCGAGTTTCATC	3630
Db			
QY	61	gtgccccgcagtgagtgaggggccctgccatccgagtgctctagccgcctggggcacc	120
Db			
QY	3631	GTGCCCCGCAGTAGTGAGGAGGGCCCTGCCATCCGAGTGCTAGCCGCTGGGSCACCCA	3690
Db			
QY	121	gttcgctacgtgtgatctcacacacagcgcgcgagctcttcgcaacgcccgcgacctctgt	180
Db			
QY	3691	GTTTGGCTACGTGGTGATCTCACACAGCCGGCAGCTTCCTGCAACAGCCCGGACTCTGT	3750
Db			
QY	181	gaacagcagccgcgcaatgtgcagcattacacaaagaatgagctgggctggtgcgatga	240
Db			
QY	3751	GAACAGCAGSCCCGCAATGTGCAGCAATTACCAAGAATGAGCTGGGCTGGTGGCGATGTA	3810
Db			
QY	241	gcctacaa	248
Db			
QY	3811	GCCTACAA	3818
Db			

RESULT	7
LOCUS	AF154114
DEFINITION	AF154114 630 bp mRNA Rattus norvegicus peptidoglycan recognition protein PGPR (Pgpr)
ACCESSION	AF154114
VERSION	AF154114
KEYWORDS	AF154114.1 GI:8132325
SOURCE	Norway rat.
ORGANISM	Rattus norvegicus
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 630)
Rehman,A., Teodecki,E.E. and Krueger,J.M.
Direct Submission
Submitted (25-MAY-1999) Veterinary Comparative Anatomy,
Pharmacology, and Physiology, Washington State University, P.O. Box
646520, Pullman, WA 99164, USA
Location/Qualifiers
1. .630
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/tissue_type="spleen"
1. .630
/genes="Pgrp"
20. .571
/genes="Pgrp"
/note="similar to Homo sapiens and Mus musculus
peptidoglycan recognition protein; similar to Trichoplusia
ni and Mus musculus TAG7"
/codon_start=1
/product="peptidoglycan recognition protein PGRP"
/protein_id="AAF73252.1"
/db_xref="GI:8132326"
/translation="MLFAWAPFALLGLADSCCFVVPVPRSEKALPSECGLKKPVRY
VWISHPAGSPSCSDCEQARNVOLYOMKOLGWCDAVNFLLIGDHVYEGRGWTK
GDHTGLPWPMSIGITFMGDYSRVPAPKRALRALNLKCGVSEGLRSNVEYVKGRHD
VQSTLSPGDLIYEIFIISWDHYRE"
BASE COUNT 134 a 193 c 166 g 137 t
ORIGIN

Query Match 7.3%; Score 40; DB 10; Length 630;
Best Local Similarity 100.0%; Pred. No. 9.4e-12;
Matches 40; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 124 cgctacgtgggtgattcacacacagccgcgagctctctgca 163
|||||
Db 146 CGCTACGTGGTGATCTCACACAGCCGCGCAGCTTCGTGCA 185
|||||

RESULT 8
CDRI31676
LOCUS 700 bp mRNA MAM 21-DEC-2000
DEFINITION Camelus dromedarius mRNA for peptidoglycan recognition protein.
ACCESSION AJ131676
VERSION AJ131676.1 GI:11990123
KEYWORDS peptidoglycan recognition protein.
SOURCE Arabian camel.
ORGANISM Camelus dromedarius
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Tylopoda; Camelidae; Camelus.
REFERENCE 1 (bases 1 to 700)
AUTHORS Kappeler,S.R., Farah,Z. and Puhan,Z.
TITLE Milk as a Source of Camel (Camelus dromedarius) Peptidoglycan
Recognition Protein
JOURNAL unpublished
REFERENCE 2 (bases 1 to 700)
AUTHORS Kappeler,S.R.
TITLE Direct Submission
JOURNAL Submitted (21-DEC-1998) Kappeler S.R., Institute of Food Science,
Swiss Federal Institute of Technology, LFO F76, 8092 Zurich,
SWITZERLAND
Location/Qualifiers
1. .700
/organism="Camelus dromedarius"
/sub_species="Somali breed"
/db_xref="taxon:9838"
/tissue_type="lactating mammary gland"
25. .606
/codon_start=1
/product="peptidoglycan recognition protein"

CDS

/protein_id="CAC19553.1"
/db_xref="GI:11990124"
/translation="MTRHCVLVWALLLSGAAREDPACGSIVPREWRALASEC
RERLRPRVYVVSHTAGSCDTPASCAQONVOSYHVRNLGWCDCVGNFLIGEDGL
VYEGRWNIKGAKHTWNPISIGISFMGNMNRVPPRALRAAQNLLACGVALGALR
SNYEKGHRDVTPLSPGDRLYEIIQTWSHYRA"
sig_peptide 25..87
mat_peptide 88..603
/product="peptidoglycan recognition protein"
polya_signal 678..683
BASE COUNT 130 a 247 c 201 g 122 t
ORIGIN

Query Match 4.7%; Score 26; DB 4; Length 700;
Best Local Similarity 100.0%; Pred. No. 0.0014;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 280 gaagccgagcgtggaacatcaagg 305
|||||
Db 337 GAAGCGGAGCTGGAACATCAAGG 362

RESULT 9
AF076483 690 bp mRNA PRI 15-AUG-1998
LOCUS Homo sapiens peptidoglycan recognition protein precursor (PGRP)
DEFINITION mRNA, complete cds.
ACCESSION AF076483
VERSION AF076483.1 GI:3342532
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 690)
AUTHORS Kang, D., Liu, G., Lundstrom, A., Gellius, E. and Steiner, H.
TITLE A peptidoglycan recognition protein in innate immunity conserved
from insects to humans
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 95 (17), 10078-10082 (1998)
MEDLINE 98374308
REFERENCE 2 (bases 1 to 690)
AUTHORS Kang, D., Liu, G., Lundstrom, A., Gellius, E. and Steiner, H.
TITLE Direct Submission
JOURNAL Submitted (08-JUL-1998) Microbiology, Stockholm University,
Stockholm S-106 91, Sweden
FEATURES
source Location/Qualifiers
1..690
/organism="Homo sapiens"
/db_xref="taxon:9606"
/tissue_type="bone marrow"
1..690
/gene="PGRP"
45..635
/gene="PGRP"
/codon_start=1
/product="peptidoglycan recognition protein precursor"
/protein_id="AAC31822.1"
/db_xref="GI:3342533"
/translation="MSRRSMLLAWALPSLLRLGAAQTEDPACCSPIVPRNEWKALAS
ECAQHLSPLRYVVSHTAGSSCTPASCOQARNVOHYHMKTLGWCDCVGNFLIGED
GLVYEGRWNIKGAKHTWNPISIGISFMGNMNRVPPRALRAAQNLLACGVALGALR
LRSNYVLKGRDVTPLSPGNQLYHLIQNPHYRSP"
45..104
/gene="PGRP"
/note="putative"
105..632
/gene="PGRP"
/product="peptidoglycan recognition protein"
BASE COUNT 129 a 250 c 188 g 123 t
ORIGIN

Query Match 4.0%; Score 22; DB 9; Length 690;
Best Local Similarity 100.0%; Pred. No. 0.31;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 461 tcaaaagcacccggatgtgca 482
|||||
Db 544 TCAAAGGACACCGGATGTGCA 565

RESULT 10
AF242517 724 bp mRNA PRI 15-AUG-2000
LOCUS Homo sapiens hypothetical protein SBB168 mRNA, complete cds.
DEFINITION AF242517
ACCESSION AF242517
VERSION AF242517.1 GI:9802032
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 724)
AUTHORS Wan, T., Zhang, W. and Cao, X.
TITLE Direct Submission
JOURNAL Submitted (08-MAR-2000) Department of Immunology, Second Military
Medical University & Shanghai Brilliance Biotechnology Institute,
800 Xiangyin Rd., Shanghai 200433, P.R. China
FEATURES
source Location/Qualifiers
1..724
/organism="Homo sapiens"
/db_xref="taxon:9606"
30..620
/codon_start=1
/product="hypothetical protein SBB168"
/protein_id="AAF99598.1"
/db_xref="GI:9802033"
/translation="MSRRSMLLAWALPSLLRLGAAQTEDPACCSPIVPRNEWKALAS
ECAQHLSPLRYVVSHTAGSSCTPASCOQARNVOHYHMKTLGWCDCVGNFLIGED
GLVYEGRWNIKGAKHTWNPISIGISFMGNMNRVPPRALRAAQNLLACGVALGALR
LRSNYVLKGRDVTPLSPGNQLYHLIQNPHYRSP"
BASE COUNT 161 a 251 c 186 g 126 t
ORIGIN

Query Match 4.0%; Score 22; DB 9; Length 724;
Best Local Similarity 100.0%; Pred. No. 0.31;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 461 tcaaaagcacccggatgtgca 482
|||||
Db 529 TCAAAGGACACCGGATGTGCA 550

RESULT 11
AC007785/c 166500 bp DNA PRI 11-JUN-1999
LOCUS Homo sapiens chromosome 19, BAC 282485 (CIT-B-344H19), complete
DEFINITION sequence.
ACCESSION AC007785
VERSION AC007785.1 GI:5042403
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 166500)
AUTHORS Lamerdin, J.E., McCreedy, P.M., Richardson, P., Sakalidis, G.,
Burkhardt-Schultz, K., Gordon, L., Scott, D., Johnson, G., Stilwagen, S.,
Phan, H., Velasco, N., Do, L., Regala, W., Terry, A., Danganan, L.,
Erler, A., Christensen, M., Georgescu, A., Avila, J., Attix, C.,
Andreise, T., Amico-Keller, G., Coefield, J., Duarte, S., Lucas, S.,
Bruce, R., Thomas, P., Quan, G., Kronmiller, B., Arellano, A.,
Sanders, C., Ow, D., Nolan, M., Trong, S., Kobayashi, A., Olsen, A.S. and
Carrano, A.V.


```

repeat_region complement(11065..11191)
                /rpt_family="AluJ"
repeat_region 11322..11619
                /rpt_family="AluSg"
repeat_region 11775..11814
                /rpt_family="AT-rich"
repeat_region complement(11864..11885)
                /rpt_family="AT-rich"
repeat_region complement(11914..11976)
                /rpt_family="AluS"
repeat_region complement(12055..12088)
                /rpt_family="AT-rich"
repeat_region complement(12186..12327)
                /rpt_family="L1M1"
repeat_region complement(12334..12384)
                /rpt_family="LINE2"
misc_feature 12784..12994
                /note="predicted exon, program: grail2exons human_1.3,
                frame: 1, quality: excellent, score: 92.000"
repeat_region 12847..12908
                /rpt_family="GC-rich"
misc_feature 12895..12994
                /note="DDS similarity to AA765166 nz79c05.s1 NCI_CGAP_GCB1
                Homo sapiens cDNA clone IMAGE:1301672; (222..123); 97%
                identity."
repeat_region 13002..13032
                /rpt_family="GC-rich"
misc_feature 13080..13162
                /note="predicted exon, program: grail2exons human_1.3,
                frame: 2, quality: excellent, score: 100.000-DPS
                similarity to AA765166 nz79c05.s1 NCI_CGAP_GCB1 Homo
                sapiens cDNA clone IMAGE:1301672; (122..48); 100%
                identity."
repeat_region complement(13799..13954)
                /rpt_family="GC-rich"

Query Match 4.0%; Score 22; DB 9; Length 166500;
Best Local Similarity 100.0%; Pred. No. 0.32;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 461 tcaagagcacccggatgtgca 482
    |||||
Db 17064 TCRAAGGACCCGGATGTGCA 17043

RESULT 12
AC091505 AC091505 188095 bp DNA MAM 18-AUG-2001
LOCUS Sus scrofa clone RP44-331G21, complete sequence.
ACCESSION AC091505
VERSION AC091505.2 GI:15209166
KEYWORDS HTG.
SOURCE pig.
ORGANISM Sus scrofa
          Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.

REFERENCE 1 (bases 1 to 188095)
AUTHORS Akhter,N., Ayele,K., Beckstrom-Sternberg,S.M., Benjamin,B.,
          Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C., Brooks,S.,
          Dietrich,N.L., Granite,S., Guan,X., Gupta,J., Haghighi,P.,
          Ho,S.-L., Idol,J.R., Karlins,E., Laric,P., Lee-Lin,S.-Q.,
          Legaspi,R., Maduro,Q.L., Maduro,V.B., Masiello,C., Mastrian,S.D.,
          McCloskey,J.C., McDowell,J., Pearson,R., Prasad,A., Shevchenko,Y.,
          Stantripop,S., Thomas,J.W., Thomas,P.J., Touchman,J.W.,
          Tsurgoeon,C., Vogt,J.L., Walker,M.A., Wetherby,K.D., Wiggins,L.,
          Zhang,L.-H. and Green,E.D.
          NISC Comparative Sequencing Initiative
          Unpublished
REFERENCE 2 (bases 1 to 188095)
AUTHORS Green,E.D.
TITLE Direct Submission
JOURNAL Submitted (27-APR-2001) NIH Intramural Sequencing Center, 8717
          Grovemont Circle, Gaithersburg, MD 20877, USA

```

```

REFERENCE 3 (bases 1 to 188095)
AUTHORS Green,E.D.
TITLE Direct Submission
JOURNAL Submitted (18-AUG-2001) NIH Intramural Sequencing Center, 8717
          Grovemont Circle, Gaithersburg, MD 20877, USA
COMMENT On Aug 18, 2001 this sequence version replaced gi:13811903.
          ----- Genome Center
          Center: NIH Intramural Sequencing Center
          Center code: NISC
          Web site: http://www.nisc.nih.gov
          Contact: nisc.mouse@nih.gov
          ----- Project Information
          ----- project name: cdd
          Center project name: cdd
          Center clone name: 331G21

This sequence was finished as follows unless otherwise noted:
all regions were double-stranded, sequenced with an
alternate chemistry, or covered by high quality data
(i.e., phred quality >= 30); an attempt was made to resolve
all sequencing problems, such as compressions and repeats;
all regions were covered by at least one plasmid subclone
or more than one M13 subclone; and the assembly was confirmed
by restriction digest.

CLONE LENGTH: This sequence represents the entire insert of
this clone unless otherwise noted. If there are overlapping
clones, the overlaps are noted in the beginning and end of
the Features section.

FEATURES             Location/Qualifiers
     source            1..188095
                        /organism="Sus scrofa"
                        /db_xref="taxon:9823"
                        /clone="RP44-331G21"
                        /clone_lib="RP44"
                        26986..27081
                        /note="single clone coverage"
     misc_feature      48857..48891
                        /note="single clone coverage"
     misc_feature      58401..58456
                        /note="single clone coverage"
     misc_feature      65003..65435
                        /note="single clone coverage"
BASE COUNT 43182 a 54602 c 51545 g 38766 t
ORIGIN

Query Match 3.8%; Score 21; DB 4; Length 188095;
Best Local Similarity 100.0%; Pred. No. 1.2;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 154 agcttctgcaacagccggac 174
    |||||
Db 64633 AGCTTTCACACAGCCCGGAC 64653

RESULT 13
RNCAS75/c RNCAS75 606 bp mRNA ROD 24-FEB-1998
LOCUS Rattus norvegicus mRNA for calpastatin, clone RNCAS75.
DEFINITION Y13591
ACCESSION Y13591
VERSION Y13591.1 GI:2765345
KEYWORDS calpastatin; CAST gene.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
          Rattus.
          1 (bases 1 to 606)
REFERENCE De Tullio,R.
AUTHORS Direct Submission
TITLE Submitted (19-MAY-1997) R. De Tullio, University of Genova, 16132,
          Institute of Biochemistry, Viale Benedetto XV, 1, Genova, ITALY

```


PS Claim 3; Fig 1; 138pp; English.

This sequence encodes the murine tag7 of the invention. Cells containing the tag7 DNA sequence are used to express recombinant tag7. Tag7 is used to produce and purify antibodies to inhibit growth of mammalian tumours, especially for treating carcinoma (of liver, ovary, breast, cervix, lung, prostate, colon/rectum, bladder, testis, stomach, pancreas, mouth, head and neck, squamous cell carcinoma or teratocarcinoma), sarcoma (Kaposi's, osteo- or fibro-sarcomas), melanoma or leukaemia; and as a molecular weight marker. The tag7 polypeptide inhibits tumour growth and induces apoptosis. The tag7 coding sequences are also useful as probes for gene mapping and detection of tag7 gene expression, and as primers. Antibodies against tag7 are used as reagents for detecting tag7; as an antagonist of tag7; for isolating tag7 and therapeutically to inhibit or delay tumour metastasis.

SQ Sequence 549 BP; 117 A; 161 C; 151 G; 120 T; 0 other;

Query Match	100.0%	Score 549;	DB 20;	Length 549;
Best Local Similarity	100.0%	Pred. No. 8.7e-156;		
Matches 549; Conservative	0;	Mismatches 0;	Indels 0;	Gaps 0;

Qy	1	atgtgttgcctgtgctctcttgcctcctcctggtctggaacctctgcagtttccate	60
Db	1	atgtgttgcctgtgctctcttgcctcctcctggtctggaacctctgcagtttccate	60
Qy	61	gtccccgcagtgtagtgaggcccttgccatccagtgctcttagccgctggggcaccca	120
Db	61	gtccccgcagtgtagtgaggcccttgccatccagtgctcttagccgctggggcaccca	120
Qy	121	gtctgctacgtgtagtattcacacagccggcagctcttgcaacagccggactcctgt	180
Db	121	gtctgctacgtgtagtattcacacagccggcagctcttgcaacagccggactcctgt	180
Qy	181	gaacagagcccgcaatgtgcagcattaccacaagaatgaactggctggtgcgatgta	240
Db	181	gaacagagcccgcaatgtgcagcattaccacaagaatgaactggctggtgcgatgta	240
Qy	241	gctcaaaccttctattggagagcggtcattctctatgaaggccgaggtcggaacatc	300
Db	241	gctcaaaccttctattggagagcggtcattctctatgaaggccgaggtcggaacatc	300
Qy	301	aaggggtgaccacacagggcccattctggaaatcccatgctctattggcatacacttoatgggg	360
Db	301	aaggggtgaccacacagggcccattctggaaatcccatgctctattggcatacacttoatgggg	360
Qy	361	aacttcattggacgggtaccgcgaagcgggcccctccgtgctgccctaaactctctggaa	420
Db	361	aacttcattggacgggtaccgcgaagcgggcccctccgtgctgccctaaactctctggaa	420
Qy	421	tgtgagggtgctcgggggttctcgagatcccaactatgaagtcaaggacacccggagtgtg	480
Db	421	tgtgagggtgctcgggggttctcgagatcccaactatgaagtcaaggacacccggagtgtg	480
Qy	481	caaaagcactctctccagggtgaccaactctatcaggtatcccaagcctgggaactac	540
Db	481	caaaagcactctctccagggtgaccaactctatcaggtatcccaagcctgggaactac	540
Qy	541	cgagagtga	549
Db	541	cgagagtga	549

RESULT	2	
AAT78510		
ID	AAT78510	standard; cDNA; 677 BP.
XX		
XX		
AC	AAT78510;	
XX		
18-FEB-1998		(first entry)
Dr		
XX		
XX		
EE		Murine granulocyte peptide A precursor cDNA.

XX	Antimicrobial peptide; antibiotic; antibacterial; antifungal;
KW	fungicide; antiprotozoa; protozoacide; antiviral; virucide;
KW	murine granulocyte peptide A; MGP-A; preservative; sepsis;
KW	endotoxaemia; mouse; ss.
XX	
OS	Mus musculus.
XX	
FH	Key
EH	36..581
FT	Location/Qualifiers
FT	/tag= a
FT	/trans_except= (pos:561..563, aa:Trp)
FT	543..578
FT	/tag= b
FT	/product= MGP-A
XX	
PN	WO9729765-A1.
XX	
XX	
PD	21-AUG-1997.
XX	
PF	13-FEB-1997; 97WO-US02218.
XX	
PR	16-FEB-1996; 96US-0011834.
XX	
PA	(REGC) UNIV CALIFORNIA.
XX	
PI	Selsted ME;
XX	
DR	WPI: 1997-424753/39.
DR	P-PSDB; AAW23723.
XX	
PT	Antimicrobial (poly)peptide(s) from bovine and murine granulocytes
PT	useful therapeutically, as preservatives for food, in water
PT	treatment and in agriculture
XX	
PS	Claim 14; Fig 5; 56pp; English.
XX	
CC	This cDNA sequence encodes the precursor (see AAW23723) of a novel,
CC	claimed antimicrobial peptide from bovine neutrophils, designated
CC	murine granulocyte peptide A or MGP-A (see AAW23725). It was
CC	isolated from murine bone marrow cDNA using primers based on
CC	bovine BGP-A cDNA (see AAW78509). The encoded MGP-A precursor
CC	comprises signal peptide and propeptide sequences followed by 12
CC	C-terminal amino acids corresponding to mature MGP-A. MGP-A and
CC	BGP-A (see AAW23724), exhibit activity against Gram-positive and
CC	Gram-negative bacteria, fungi and viruses, specifically
CC	Staphylococcus aureus, Escherichia coli, Candida albicans,
CC	Salmonella typhimurium and C. neoformans (claimed). They can be
CC	used in human or veterinary medicine (particularly to treat
CC	diseases associated with lipopolysaccharides, e.g. sepsis and
CC	endotoxaemia) or as preservatives in food products or in water
CC	supplies (claimed). They can also be applied to crops to increase
CC	post-harvest spoilage or expressed in transgenic plants to increase
CC	their disease resistance. They have low immunogenicity.
XX	
SO	Sequence 677 BP; 146 A; 217 C; 171 G; 143 T; 0 other:

Query Match.	95.38;	Score 523.4;	DB 18;	Length 677;
Best Local Similarity	98.48;	pred. No. 4.9e-148;		
Matches 540;	Conservative	0;	Mismatches 6;	Indels 3;
				Gaps

Qy	1	atgtgttttgcctgtgtctctctcttgcctcctggtgtggtggaacctctcgtcagtttcac	60
Db	36	atgtgttttgcctgtgtctctctcttgcctcctcctggtgtggtggaacctctcgtcagtttcac <td>95</td>	95
Qy	61	gtccccgcagtgagtgagggggcctctgcctacgcagtgctctctagccgcctctggggacccca <td>120</td>	120
Db	96	gtgccccgcagtgagtgagggggcctctgcctacgcagtgctctctagccgcctctggggacccca <td>155</td>	155
Qy	121	gttcgctacgtggtgatctctcacacagccgcgcagctctctgcaacagccgcgactcctgt <td>180</td>	180
Db	156	gttcgctacgtggtgatctctcacacagccgcgcagctctctgcaacagccgcgactcctgt <td>215</td>	215